

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4 p3_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 9, 2003, 18:47:19; Search time 1241 Seconds

(without alignments) 8207.882 million cell updates/sec

Title: US-09-880-711-328

Perfect score: 350
Sequence: 1 atgagctcaccagctaccctcga.....ctgagagcttcaataaaaa 350

Scoring table: IDENTITY_NUCDX
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4105280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_hgt: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pal: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_of: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_hgt_hum: *
31: em_hgt_inv: *
32: em_hgt_other: *
33: em_hgt_mus: *
34: em_hgt_pln: *
35: em_hgt_rtd: *
36: em_hgt_mam: *
37: em_hgt_vrt: *
38: em_sy: *
39: em_hgtgo_hum: *
40: em_hgtgo_mus: *
41: em_hgtgo_other: *

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query				DB ID	Description
	Score	Match	Length	DB		
1	201.2	57.5	1315	5	BC003608	BC003608 Homo sapi
2	199.6	57.0	1326	9	D84109	D84109 Homo sapien
3	171.6	49.0	2406	10	BC011288	BC011288 Mus muscu
4	169.4	48.4	1587	10	AF148511	AF148511 Mus muscu
5	167.8	47.5	1124	10	BC030397	BC030397 Mus muscu
6	150.4	43.0	209598	2	AC099473	AC099473 Rattus no
7	142.8	40.8	1594	9	D84110	D84110 Homo sapien
8	141.8	40.5	778	9	D84111	D84111 Homo sapien
9	141.8	40.5	1388	9	D84107	D84107 Homo sapien
10	141.4	40.4	61962	2	AC123616	AC123616 Mus muscu
11	141.4	40.4	262432	2	AC127551	AC127551 Mus muscu
12	124.2	35.5	1995	5	AK057533	AK057533 Homo sapi
13	124.2	35.5	126753	2	AF252826	AF252826 Homo sapi
14	124.2	35.5	127677	9	AC102945	AC102945 Homo sapi
15	124.2	35.5	155926	9	AC011204	AC011204 Homo sapi
16	117.8	33.7	1474	5	D84108	D84108 Homo sapien
17	95.2	27.2	113320	2	AC122975	AC122975 Rattus no
18	72	20.6	93263	2	AC103070	AC103070 Rattus no
19	64.8	18.5	60155	5	AC083883	AC083883 Homo sapi
20	63.6	18.2	92143	9	HS608E8	AL022343 Human DNA
21	63.4	18.1	37239	9	AC096549	AC096549 Homo sapi
22	62.6	17.9	3293	8	AF073629	AF073629 Zea mays
23	62.6	17.9	174192	2	AC110534	AC110534 Mus muscu
24	62.6	17.9	235593	2	AL772296	AL772296 Mus muscu
25	62.6	17.9	293387	2	AL807802	AL807802 Mus muscu
26	62.2	17.8	265938	2	AC127247	AC127247 Mus muscu
27	61.6	17.6	125020	9	AF429315	AF429315 Homo sapi
28	61.4	17.5	158836	2	AC106126	AC106126 Rattus no
29	61.4	17.5	170523	2	AC053526	AC053526 Homo sapi
30	61.4	17.5	170821	10	AL645626	AL645626 Mouse DNA
31	61.4	17.5	177782	9	AC117532	AC117532 Homo sapi
32	61.4	17.5	178105	9	AC073057	AC073057 Homo sapi
33	61.4	17.5	189648	2	AC118994	AC118994 Rattus no
34	61.2	17.5	152019	2	AC129992	AC129992 Rattus no
35	61	17.4	179714	9	AC092959	AC092959 Homo sapi
36	61	17.4	184696	9	AC112507	AC112507 Homo sapi
37	60.8	17.4	7218	6	166494	166494 Sequence 14
38	60.8	17.4	135721	2	AC119606	AC119606 Rattus no
39	60.5	17.3	39601	9	AC005581	AC005581 Homo sapi
40	60.6	17.3	131903	9	HS360E18	Z82203 Human DNA s
41	60.6	17.3	139843	2	AL158145	AL158145 Homo sapi
42	60.6	17.3	160069	2	AC023329	AC023329 Homo sapi
43	60.2	17.2	40468	3	AF040654	AF040654 Caenorhab
44	60.2	17.2	139258	9	HSDA60H16	AL050305 Human DNA
45	60	17.1	2481	9	HOMNAK1	D49728 Human NAK1

ALIGNMENTS

RESULT 1
LOCUS BC003608 1315 bp mRNA linear 12-JUL-2001
DEFINITION Homo sapiens, KNA-binding protein gene with multiple splicing,
clone MGC:1257 IMAGE:3505124, mRNA, complete cds.
ACCESSION BC003608
VERSION BC003608.1 GI:13177643
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1315)
AUTHORS Strausberg, R.
TITLE Direct Submission

JOURNAL

Submitted (26-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-5590, USA

REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNI)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
contact: amadansystemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia Greene, Mark Kettelman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNI at: <http://image.llni.gov>
Series: IRAL Plate: 7 Row: P Column: 3
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 5803140.

FEATURES

source

1. 1315
/organism="Homo sapiens"
/db_xref="LocusID:11030"
/db_xref="taxon:9606"
/clone="MGC:1257 IMAGE:3505124"
/tissue_type="Placenta, choriocarcinoma"
/clone_id="NIH.MGC-21"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
543..1202
/codon_start=1
/product="RNA-binding protein gene with multiple splicing"
/protein_id="AAH03608.1"
/db_xref="GI:13177644"
/translation="MNNGGKAEKENTPSEANLOEEVRLTVSGCLPLDIPRELYLLE
RPFKGEISLILKLSKOPVGFVSFDSSEAEAKNALINGIRFDEIPQTLREFAKAN
TKMKNKLVGTPNPSTPLPNTVPOFIAREPELVLPALYSSPEVMAPIPLYPALAP
ALPPPAFTYPASLHAOCFSPKAKPNTVPCPLIQIRFVSGNVEVTYQPTAQQRELPL
C"

CDS

BASE COUNT 265 a 483 c 290 g 277 t
ORIGIN
Query Match 57.5%; Score 201.2; DB 9; Length 1315;
Best Local Similarity 71.4%; Pred. No. 1.8e-23;
Matches 250; Conservative 29; Mismatches 58; Indels 13; Gaps 3;

1 ATGAGCTCAGTACGACCTTACCCCGAGTACGCTGAGTGTGGCCCGCTACCCCTC 60
|||||
940 ATGAGCTCAGTACGACCTTACCCCGAGTACGCTGAGTGTGGCCCGCTACCCCTC 99
61 TGTACCCAGGAGTGTAGCCGCTGCTTCCCTCCTCCGCGCTTACCTACCCGCTT 120
|||||
1000 TGTACCCAGGAGTGTAGCCGCTGCTTCCCTCCTCCGCGCTTACCTACCCGCTT 1056
121 CACTGATGCGCAGAGCAGCTGATATGATGCTAGAGNNNNACNCTGACNNNTGAGANAC 180
|||||
1057 CACTGATGCGCAGAGCAGCTGATATGATGCTAGAGNNNNACNCTGACNNNTGAGANAC 1116
181 TNCATGAGCAGATGATGCTGATGCTGAGTGTGGCCCGCTACCCCTACCCCTC 240
|||||
1117 TTCTCAGCAAAAT--TAAATTTGCTCTGGGAAATGTGTTGTAACATACCAACCTACG 1173
241 CAGACCAACNNAAGGAGCTGNNGACAGATGATATTANNACCNACCTACCTATNCCTT 300
|||||
1174 CAGACCAACNNAAGGAGCTGNNGACAGATGATATTANNACCNACCTACCTATNCCTT 1226
301 TTGCAANANMAACTTCTTTGANAATTTTGGAGATTTCAATTAATAA 350
|||||
1227 CCTTTCACAAATACTATTCTTGTGACGACCTTTGAGAGATTCAATTAATAA 1276

RESULT 2

DB4109 1326 bp. mRNA linear PRI 14-APR-2000
LOCUS Homo sapiens mRNA for RBP-MS/Type 3, complete cds.
DEFINITION DB4109.1 GI:1669550
ACCESSION
VERSION
KEYWORDS RBP-MS; RBP-MS/Type 3; alternative splicing; Werner syndrome.
SOURCE Homo sapiens cell line: embryonic carcinoma, NEC14 CDNA to mRNA.
ORGANISM Homo sapiens

REFERENCE

1 (sites)
Shimamoto, A., Kitao, S., Ichikawa, K., Suzuki, N., Yamabe, Y., Imamura, O., Tokutake, Y., Satoh, M., Matsunoto, T., Kurumitsu, J., Katoaka, H., Sugawara, K., Sugawara, M., Sugimoto, M., Goto, M. and Furukuchi, Y.
A unique human gene that spans over 230 kb in the human chromosome 8p11-12 and codes multiple family proteins sharing RNA-binding motifs

TITLE

JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (21-MAR-1996) Akira Shimamoto, AGNE Research Institute Co., Ltd., 200 Kajiwara, Kamakura, Kanagawa 247, Japan
(E-mail: akirashimo@po.ijnet.or.jp, Tel: 0467-46-4971, Fax: 0467-48-6595)

FEATURES

source

1. 1326
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8p11.2-12.1"
/cell_line="embryonic carcinoma, NEC14"
/gene="RBP-MS"
1..647
/gene="RBP-MS"
/number=1
1..566
/gene="RBP-MS"
567..1226
/gene="RBP-MS"
/standard_name="RNA-binding protein gene with multiple splicing"
/note="alternative splicing (see also D18407-D84111)"
/codon_start=1
/product="RBP-MS/Type 3"
/protein_id="BAI12227.1"
/db_xref="GI:1669551"
/translation="MNNGGKAEKENTPSEANLOEEVRLTVSGCLPLDIPRELYLLE
RPFKGEISLILKLSKOPVGFVSFDSSEAEAKNALINGIRFDEIPQTLREFAKAN
TKMKNKLVGTPNPSTPLPNTVPOFIAREPELVLPALYSSPEVMAPIPLYPALAP
ALPPPAFTYPASLHAOCFSPKAKPNTVPCPLIQIRFVSGNVEVTYQPTAQQRELPL
C"
648..706
/gene="RBP-MS"
/number=2
707..756
/gene="RBP-MS"
/number=3
757..812
/gene="RBP-MS"
/number=4
813..963
/gene="RBP-MS"
/number=5
964..1094
/gene="RBP-MS"
/number=7
1095..1326
/gene="RBP-MS"

exon

	/number=10	
BASE COUNT	745 a	497 c
ORIGIN	298 g	286 t

ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 209598)
REFERENCE.
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-ouman, F.R., Allen, C.,
AUTHORS

exon	648.	706	/gene="RBP-MS"
exon	707.	756	/number=2
exon	757.	812	/gene="RBP-MS"
exon	757.	812	/number=3
exon	813.	963	/gene="RBP-MS"
exon	964.	1094	/number=5
exon	1095.	1164	/gene="RBP-MS"
exon	1165.	1370	/number=12
exon	1371.	1471	/gene="RBP-MS"
exon	1472.	1594	/number=14
exon	1594.	1594	/gene="RBP-MS"
exon	1594.	1594	/number=15

BASE COUNT	313 a	563 c	373 g	345 t
ORIGIN				
Query Match	40.8%	Score 142.8;	DB 9;	Length 1594;
Best Local Similarity	64.9%;	Pred. No. 4.1e-13;		
Matches 187;	Conservative 24;	Mismatches 72;	Indels 5;	Gaps 2;

exon	648.	706	/gene="RBP-MS"
exon	707.	756	/number=2
exon	757.	812	/gene="RBP-MS"
exon	757.	812	/number=3
exon	813.	963	/gene="RBP-MS"
exon	964.	1094	/number=5
exon	1095.	1164	/gene="RBP-MS"
exon	1165.	1370	/number=12
exon	1371.	1471	/gene="RBP-MS"
exon	1472.	1594	/number=14
exon	1594.	1594	/gene="RBP-MS"
exon	1594.	1594	/number=15

exon	648.	706	/gene="RBP-MS"
exon	707.	756	/number=2
exon	757.	812	/gene="RBP-MS"
exon	757.	812	/number=3
exon	813.	963	/gene="RBP-MS"
exon	964.	1094	/number=5
exon	1095.	1164	/gene="RBP-MS"
exon	1165.	1370	/number=12
exon	1371.	1471	/gene="RBP-MS"
exon	1472.	1594	/number=14
exon	1594.	1594	/gene="RBP-MS"
exon	1594.	1594	/number=15

exon	648.	706	/gene="RBP-MS"
exon	707.	756	/number=2
exon	757.	812	/gene="RBP-MS"
exon	757.	812	/number=3
exon	813.	963	/gene="RBP-MS"
exon	964.	1094	/number=5
exon	1095.	1164	/gene="RBP-MS"
exon	1165.	1370	/number=12
exon	1371.	1471	/gene="RBP-MS"
exon	1472.	1594	/number=14
exon	1594.	1594	/gene="RBP-MS"
exon	1594.	1594	/number=15

exon	648.	706	/gene="RBP-MS"
exon	707.	756	/number=2
exon	757.	812	/gene="RBP-MS"
exon	757.	812	/number=3
exon	813.	963	/gene="RBP-MS"
exon	964.	1094	/number=5
exon	1095.	1164	/gene="RBP-MS"
exon	1165.	1370	/number=12
exon	1371.	1471	/gene="RBP-MS"
exon	1472.	1594	/number=14
exon	1594.	1594	/gene="RBP-MS"
exon	1594.	1594	/number=15

exon	648.	706	/gene="RBP-MS"
exon	707.	756	/number=2
exon	757.	812	/gene="RBP-MS"
exon	757.	812	/number=3
exon	813.	963	/gene="RBP-MS"
exon	964.	1094	/number=5
exon	1095.	1164	/gene="RBP-MS"
exon	1165.	1370	/number=12
exon	1371.	1471	/gene="RBP-MS"
exon	1472.	1594	/number=14
exon	1594.	1594	/gene="RBP-MS"
exon	1594.	1594	/number=15

exon	648.	706	/gene="RBP-MS"
exon	707.	756	/number=2
exon	757.	812	/gene="RBP-MS"
exon	757.	812	/number=3
exon	813.	963	/gene="RBP-MS"
exon	964.	1094	/number=5
exon	1095.	1164	/gene="RBP-MS"
exon	1165.	1370	/number=12
exon	1371.	1471	/gene="RBP-MS"

TITLE	A unique human gene that spans over 230 kb in the human chromosome 8p11-12 and codes multiple family proteins sharing RNA-binding motifs						
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 93 (20), 10913-10917 (1996)						
MEDLINE	97003106						
REFERENCE	2 (bases 1 to 778)						
AUTHORS	Shimamoto A.						
TITLE	Direct Submission						
JOURNAL	Submitted (21-MAR-1996) Akira Shimamoto, AGRIE Research Institute Co., Ltd.; 200 Kajiwara, Kamakura, Kanagawa 247, Japan (E-mail: akirashimo@po.lljnet.or.jp, Tel: 0467-46-4971, Fax: 0467-48-6595)						
FEATURES	Location/Qualifiers						
Source	1..778						
	/organism="Homo sapiens"						
	/db_xref="taxon:9606"						
	/chromosome="8"						
	/map="8p11.2-12.1"						
	/cell_line="HeLa"						
gene	1..778						
	/gene="RBP-MS"						
CDS	<1..298						
	/gene="RBP-MS"						
	/standard_name="RNA-binding protein gene with multiple splicing"						
	/note="alternative splicing (see also D84107-D84111)"						
	/codon_start=2						
	/product="RBP-MS/type 5"						
	/protein_id="BAA12224.1"						
	/cd_xref="GI:1659535"						
	/translation="PFRGYEGSLIKLTSPQVGFVSFSDSRSEAEAKNALNGIREDPEIPQLRLFEFAKANKRKAKNKLVGTPNPSTPIPLNIVPOFIAREPLGANEIGERO"						
exon	<1..6						
	/gene="RBP-MS"						
	/number=2						
exon	7..56						
	/gene="RBP-MS"						
	/number=3						
exon	57..112						
	/gene="RBP-MS"						
	/number=4						
exon	113..263						
	/gene="RBP-MS"						
	/number=5						
exon	264..409						
	/gene="RBP-MS"						
	/number=6						
exon	410..540						
	/gene="RBP-MS"						
	/number=7						
exon	541..610						
	/gene="RBP-MS"						
	/number=12						
exon	611..711						
	/gene="RBP-MS"						
	/number=14						
exon	712..778						
	/gene="RBP-MS"						
	/number=15						
BASE COUNT	180 a 211 c 181 g 206 t						
ORIGIN							
Query Match	40.5%; Score 141.8; DB 9; Length 778;						
Best Local Similarity	64.9%; Pred. No. 4,66-13;						
Matches 185; Conservative	24; Mismatches 72; Indels 4; Gaps 2;						
OY	1 ATGAGCTCACAGTACTACCTTTATCCCGCAGTAGCCGTGAAGTCGTGGGCCCGGTACCCTGC 60 						
DB	410 ATGAGCTCACAGTACTACCTTTATCCCGCAGTAGCCGTGAAGTCGTGGGCCCGGTACCCTGC 469 						
OY	61 TGTACCGACGGAGTTAGCCGCTGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120 						
DB	470 TGTACCGACGGAGTTAGCCGCTGCT 525 						

OY	121	CAGTCGATGCCAGAGAGCCTGTATACATGGCTAGGNNNCA [^] CNCCTGCACNNNTGGANAG	180
D6	527	CAGTCCATGCCAGAGAGCGCTGGCTCCCTCCCGACGACTACTCTCAGGGCTGGAGCT	586
OY	181	TGCTGACGACGATGATGATGCTGCTGCTGCTGGGAATGCGCTGCACACATCACTACTG	240
D6	587	CGCGTGATCTG-TGGTAGAATACTATGTTCCAGAGTGTGTATGATGGCGGCTGCACAATGTGTCTT	645
OY	241	CAGACCAG-NAAGGACGCTTNGACACAAATGATT [^] NNAACCN	285
		: : : : : : :	
D6	646	GTCGGATTATATGCAATCTTCAGTGGTGGCTACTGTTCTCTAGCT	690

RESULT 9			
LOCUS	D84107	1388 bp	mRNA linear PK1 14-Apr-2000
DEFINITION	Homo sapiens mRNA for RBP-MS/type 1, complete cds.		
ACCESSION	D84107		
VERSION	D84107.1 GI:1669546		
KEYWORDS	RBP-MS; RBP-MS/type 1; alternative splicing; Werner syndrome-		
SOURCE	Homo sapiens cell_line=embryonic carcinoma, NCI4 cDNA to mRNA.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
	Shimamoto,A., Kitao,S., Ichikawa,K., Suzuki,N., Yamabe,Y.,		
	Iinamura,O., Tokutake,Y., Satoh,M., Matsumoto,T., Kurokisu,J.,		
	Kataoka,H., Sugawara,K., Sugawara,M., Sugimoto,M., Goto,M. and		
	Furuchi,Y.		
TITLE	A unique human gene that spans over 230 kb in the human chromosome		
	8p11-12 and codes multiple family proteins sharing RNA-binding		
	motifs		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 93 (20), 10913-10917 (1996)		
MDLINE	97008106		
REFERENCE	2 (bases 1 to 1388)		
AUTHORS	Shimamoto,A.		
TITLE	Direct Submission		
JOURNAL	Submitted (21-MAR-1996) Akira Shimamoto, AGENE Research Institute		
	Co., Ltd.; 200 Kajiwara, Kamakura, Kanagawa 247, Japan		
	(E-mail:kajirashi@istio.po.tijnet.or.jp, Tel:0467-46-4971,		
	Fax:0467-48-6595)		

FEATURES	location/Qualifiers
source	1..1388
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="8"
	/map="gq11.2-12.1"
	/cell_line="embryonic carcinoma, NEC14"
	1..1388
gene	/gene="RBP-MS"
	1..647
exon	/gene="RBP-MS"
	/number=1
	1..566
5'UTR	/gene="RBP-MS"
	567..1157
CDS	/gene="RBP-MS"
	/standard_name="RNA-binding protein gene with multiple
	splicing"
	/note="alternative splicing (see also D84107-D8411)"
	/codon_start=1
	/product="RBP-MS/type 1"
	/protein_id="BAI2225.1"
	/cd_xref="GI:1669547"
	/translation="MNNQKAERENTPSEANLQEEVETFLVSLPIPRELYLP
	RPFGYSKLTKLSKDPGVFSFSDSPAAFNALNGIPDEIPQTALRFKXAN
	TKMKNLKGTGNSTPLPNVPQFIAREPELTYPYSSPEVMADPYLPALPAD
	ALPLAPITYPASLIAQBRMLPPSEATSGQSKSFQ"
exon	648..706
	/gene="RBP-MS"
	/number=2
	707..756

Query Match	40.5%;	Score 141.8;	DB 9;	Length 1388;
Best Local Similarity	64.9%;	Pred. No. 5.8e-13;		
Matches 185;	Conservative 24;	Mismatches 72;	Indels 4;	Gaps 2

RESULT	10
LOCUS	AC123616
DEFINITION	AC123616 Mus musculus clone RP23-4f1, LOW-PASS SEQUENCE SAMPLING.
ACCESSION	AC123616
VERSION	AC123616.1 GI:21307015
KEYWORDS	HTG; HTGS_PHASEO.
SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE	AUTHORS	JOURNAL	TITLE
1 (bases 1 to 61962)	Bliren, B., Linton, L., Nussbaum, C. and Lander, E.	Mus musculus, clone RP23-4F1	Unpublished
2 (bases 1 to 61962)	Bliren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Brown, A., Camarata, J., Campioniano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collimore, A., Cook, A., Cooke, P., Deatrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fato, S., Ferreira, P., Fitzgerald, M., FitzHugh, W., Gage, D., Galadagen, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand, P., Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karaats, A., Kells, C., Larocque, K.,		

Limaazares, R., Landers, T., Lehoczyk, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Majtor, J., Marquis, N., Matthews, C., McCarthy, M., McEwen, P., McKernan, K., Meldrum, J., Meneses, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Neuyen, C., Nicoli, R., Notbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, K., Riley, R., Rise, C., Rogov, P., Roman, J., Rossetti, M., Roy, A., Santos, B., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talmas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travs, N., Tirigillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zaimoun, J., Zemek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (01-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RV/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIMR
Web site: <http://www.seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: I23216
Center clone name: 4_F_1

* NOTE: this record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allow
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved. 691 692 693 694 695 696 697 698 699 700

*	1	681: contig of 681 bp in length
*	662	781: gap of 100 bp
*	782	1550: contig of 769 bp in length
*	1551	1650: gap of 100 bp
*	1651	2350: contig of 700 bp in length
*	2351	2450: gap of 100 bp
*	2451	3161: contig of 711 bp in length
*	3162	3261: gap of 100 bp
*	3262	3964: contig of 703 bp in length
*	3965	4064: gap of 100 bp
*	4065	4766: contig of 704 bp in length
*	4769	4868: gap of 100 bp
*	4869	5556: contig of 688 bp in length
*	5557	5656: gap of 100 bp
*	5657	6323: contig of 667 bp in length
*	6324	6423: gap of 100 bp
*	6424	7108: contig of 685 bp in length
*	7109	7208: gap of 100 bp
*	7209	7892: contig of 684 bp in length
*	7893	7992: gap of 100 bp
*	7993	8655: contig of 673 bp in length
*	8656	8765: gap of 100 bp
*	8766	9479: contig of 714 bp in length
*	9480	9579: gap of 100 bp
*	9580	10280: contig of 701 bp in length
*	10281	10380: gap of 100 bp
*	10381	11084: contig of 704 bp in length
*	11085	11184: gap of 100 bp
*	11185	11888: contig of 704 bp in length
*	11889	11988: gap of 100 bp
*	11989	12689: contig of 701 bp in length
*	12690	12789: gap of 100 bp
*	12790	132486: contig of 657 bp in length
*	13487	13586: gap of 100 bp

```
* 13587 14284: contig of 698 bp in length
* 14285 14384: gap of 100 bp
* 14385 15071: contig of 687 bp in length
* 15072 15171: gap of 100 bp
* 15172 15872: contig of 701 bp in length
* 15873 15972: gap of 100 bp
* 15973 16662: contig of 690 bp in length
* 16663 16762: gap of 100 bp
* 16763 17452: contig of 690 bp in length
* 17453 17552: gap of 100 bp
* 17553 18261: contig of 709 bp in length
* 18262 18361: gap of 100 bp
* 18362 19037: contig of 676 bp in length
* 19038 19137: gap of 100 bp
* 19138 19840: contig of 703 bp in length
* 19841 19940: gap of 100 bp
* 19941 20649: contig of 709 bp in length
* 20650 20749: gap of 100 bp
* 20750 21449: contig of 700 bp in length
* 21450 21549: gap of 100 bp
* 21550 22220: contig of 671 bp in length
* 22221 22320: gap of 100 bp
* 22321 23010: contig of 690 bp in length
* 23011 23110: gap of 100 bp
* 23111 23802: contig of 692 bp in length
* 23803 23902: gap of 100 bp
* 23903 24592: contig of 690 bp in length
* 24593 24692: gap of 100 bp
* 24693 25403: contig of 711 bp in length
* 25404 25503: gap of 100 bp
* 25504 26201: contig of 698 bp in length
* 26202 26301: gap of 100 bp
* 26302 27005: contig of 704 bp in length
* 27006 27105: gap of 100 bp
* 27106 27772: contig of 667 bp in length
* 27773 27872: gap of 100 bp
* 27873 28565: contig of 693 bp in length
* 28566 28665: gap of 100 bp
* 28666 29362: contig of 697 bp in length
* 29363 29462: gap of 100 bp
* 29463 30154: contig of 692 bp in length
* 30155 30254: gap of 100 bp
* 30255 30972: contig of 718 bp in length
* 30973 31072: gap of 100 bp
* 31073 31756: contig of 684 bp in length
* 31757 31856: gap of 100 bp
* 31857 32525: contig of 669 bp in length
* 32526 32625: gap of 100 bp
* 32626 33318: contig of 693 bp in length
* 33319 33418: gap of 100 bp
* 33419 34129: contig of 711 bp in length
* 34130 34229: gap of 100 bp
* 34230 34935: contig of 706 bp in length
* 34936 35035: gap of 100 bp
* 35036 35732: contig of 697 bp in length
* 35733 35832: gap of 100 bp
* 35833 36538: contig of 706 bp in length
* 36539 36638: gap of 100 bp
* 36639 37325: contig of 687 bp in length
* 37326 37425: gap of 100 bp
* 37426 38119: contig of 694 bp in length
* 38120 38219: gap of 100 bp
* 38220 38900: contig of 681 bp in length
* 38901 39000: gap of 100 bp
* 39001 39704: contig of 704 bp in length
* 39705 39804: gap of 100 bp
* 39805 40499: contig of 695 bp in length
* 40500 40599: gap of 100 bp
* 40600 41284: contig of 685 bp in length
* 41285 41384: gap of 100 bp
* 41385 42075: contig of 691 bp in length
* 42076 42175: gap of 100 bp
* 42176 42868: contig of 693 bp in length
```

```
* 42869 42968: gap of 100 bp
* 42969 43673: contig of 705 bp in length
* 43674 43773: gap of 100 bp
* 43774 44483: contig of 710 bp in length
* 44484 44583: gap of 100 bp
* 44584 45279: contig of 696 bp in length
* 45280 45379: gap of 100 bp
* 45380 46065: contig of 686 bp in length
* 46066 46165: gap of 100 bp
* 46166 46858: contig of 693 bp in length
* 46859 46958: gap of 100 bp
* 46959 47645: contig of 687 bp in length
* 47646 47745: gap of 100 bp
* 47746 48418: contig of 673 bp in length
* 48419 48518: gap of 100 bp
* 48519 49234: contig of 716 bp in length
* 49235 49334: gap of 100 bp
* 49335 50021: contig of 687 bp in length
* 50022 50121: gap of 100 bp
* 50122 50814: contig of 693 bp in length
* 50815 50914: gap of 100 bp
* 50915 51630: contig of 716 bp in length
* 51631 51730: gap of 100 bp
* 51731 52430: contig of 700 bp in length
* 52431 52530: gap of 100 bp
* 52531 53243: contig of 713 bp in length
* 53244 53343: gap of 100 bp
* 53344 54043: contig of 700 bp in length
* 54044 54143: gap of 100 bp
* 54144 54841: contig of 698 bp in length
* 54842 54941: gap of 100 bp
```

Query Match 40.4% Score 141.4: DB 2: Length 61962:

Best Local Similarity 65.8% Pred. No. 2.9e-12:

Matches 173: Conservative 21: Mismatches 66: Indels 3: Gaps 1:

OY 1 ATGAGCTCAAGTACTGTGACTTTACCCGAGTAGCCCTGAGTAGTGAGGCCCGCTACCTC 60

DB 53740 ATGAGCTCAAGTACTGTGACTTTACCCGAGTAGCCCTGAGTAGTGAGGCCCGCTACCTC 53799

OY 61 TGTACCCAGCGAGTTAGCGCTGCTCTCTCTCTGCGGCTTTACCTACCCGCTT 120

DB 53800 TGTACCCAGCGAGTTAGCGCTGCTCTCTCTCTCTGCGGCTTTACCTACCCGCTT 53859

OY 121 CACTGCATGCGGAGCGGCTGTATATAGCTCAGGNNNACNCGAGNNNTGANNAC 180

DB 53860 CACTGCATGCGGAGCGGCTGTATATAGCTCAGGNNNACNCGAGNNNTGANNAC 53919

OY 181 TNCCTNAGCAGCATGNNATGATNCGCTGCTGGAATGNGCTGNCAGATGACCTACTG 240

DB 53920 CCTTTGAGACCGCGCT---GCCCGAGAGACAGCATGTGACTTACCTACCGAGTAACTA 53975

OY 241 CAGACCANCANAGGACGCTTNG 263

DB 53977 GATTCACCTTTCAGGAGTGAATG 53999

RESULT 11

AC127551

LOCUS 262432 bp DNA linear HTG 11-AUG-2002

DEFINITION Mus musculus chromosome UNK clone RP24-547H23, WORKING DRAFT

AC127551

AC127551.5 G1:22138547

KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT;

SOURCE house mouse;

ORGANISM Mus musculus;

REFERENCE Fukuyama, M: Chordata; Craniata; Vertebrata; Euteleostomi;

1 (bases 1 to 262432)

McPherson, J.D. and Waterston, R.H.

TITLE The sequence of Mus musculus clone

JOURNAL Unpublished

```

-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
Project Information
Center project name: M_BB0547H23
-----
Summary Statistics
-----
Sequencing vector: MJ3, 0%
Sequencing vector: plasmid, 100%
Chemistry: Dye-primer ET, 0% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap, version 0.990319
Consensus quality: 258204 bases at least Q40
Consensus quality: 259283 bases at least Q30
Consensus quality: 259803 bases at least Q20

```

```
misc_feature 1.1128
               /note="assembly_name:Contig5"
misc_feature 1229.10138
               /note="assembly_name:Contig19
misc_feature 10239.16401
               /note="assembly_name:Contig20
misc_feature 16502.25546
               /note="assembly_name:Contig21
misc_feature 25647.44334
               /note="assembly_name:Contig22
misc_feature 44335.77154
               /note="assembly_name:Contig23
```

Query Match	40.4%	Score 141.4;	DB 2;	Length 262432;
Best Local Similarity	65.8%	Pred. No. 5,1e-12;		
Matches 173;	Conservative 21;	Mismatches 66;	Indels 3;	Gaps 1

RESULT 12	AK057533	LOCUS	DEFINITION	1995 bp	mRNA	linear	PRI 01-AUG-2002
		AK057533	AK057533	Homo sapiens CDNA FLJ32971 fis, clone TEST12008847.			
		AK057533.1	GI:16553225	AK057533.1	GT:16553225		
			oligo capping; fis (full insert sequence);				
			Homo sapiens testis CDNA to mRNA, clone_t1b:TEST12				
			clone:TEST12008847.				

TITLE
NEO human cDNA sequencing project

FEATURES	Location/Qualifiers
Research Association for Biotechnology (RAB): cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.): 5' - 3' end one pass sequencing: RAB HRI and Biotechnology Center, National Institute of Technology and Evaluation: clone selection for full insert sequencing: RAB and HRI.	


```
Source
1. 1995
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TEST12008847"
/issue_type="testis"
/clone_id="TEST12"
/note="cloning vector: pME18SFL3"
362..757
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB71523.1"
/db_xref="GI:1653276"
/translation="MITDLSIVIPINRKVDFPWSVEALPLSHGPRGSAVSGHNS
FIGSPFSPSGRISCFVDELPTVIGSPATVKTSEKSRSPRLRSDGAROADACGF
SCVIFPFLISCTKANKKQVYKRMVY1"

BASE COUNT      485 a      462 c      434 g      614 t
ORIGIN
Query Match      35.5%; Score 124.2; DB 9; Length 1995;
Best Local Similarity 58.7%; Pred. No. 8.6e-10;
Matches 192; Conservative 29; Mismatches 98; Indels 8; Gaps 3;

QY 1 ATGAGCCACACTACTCTGACCTTACCCAGAGCCCTGAGTGGCCCGCTACCCCTC 60
    |||||
DB 861 ATGAGCTCAGAGTGGCTGACCTTACCCAGAGCCCTGAGTGGCCCGCTACCCCTC 920

QY 61 TGTACCCACGAGGAGTTAGCCGCTGCTTCTGCTGCTGCTTTCAGCTACCCGCTT 120
    |||||
DB 921 TGTACCCACGAGGAGTTAGCCGCTGCTGCTGCTGCTTTCAGCTACCCGCTT 977

QY 121 CAGTCGATGCCAGAGGCGCTGTATACAT-GCCTAGAGNNNNACMCCTGCACNNNTGANA 179
    |||||
DB 978 CAGTCGATGCCAGGAGTAAATGATACCAATCGGAGAGACTTACTGACTTCACACCACT 1037

QY 180 CTNCTNACGACATGATNATGATGCGTCGATGCGATGCGATGCGATGCGATGCGATGCGAT 239
    |||||
DB 1038 CCTTCCAAACTGTTGTTCCCGAGAGCAGCAC---CATACAACTAACACCTATCCACTTAC 1093

QY 240 GCAGACACANAGAGGACCTTNNAGACAGAAATGATTAATNACNACTACTTATNCTT 299
    |||||
DB 1094 AGATCCACCTCTCAAGAGTATGATGATCCCTTAATATCAACGTGACACTCTTCAGACT 1153

QY 300 TTTCGAANAGNAACTTTCTTTGANA 326
    |||||
DB 1154 CAGCAGCATGCTGTGACACGATGATA 1180

RESULT 13
AF252826      126753 bp      FNA      Linear      HTG 26-JUL-2002
LOCUS        Homo sapiens chromosome 8 clone CTD-2373N4 map 8p12, 8 unordered
pieces.
ACCESSION    AF252826
VERSION      AF252826.3 GI:14329036
KEYWORDS     HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_CANCELLED.
SOURCE       Homo sapiens.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 126753)
Schlissel, M.B., Baumgart, C., Blechschmid, K., Dette, M., Jahn, N.,
Leimann, R., Menzel, U., Polley, A., Reichwald, K., Schudy, A.,
Siddiqui, R., Taudien, S., Wen, G., Rosenthal, A. and Platzer, M.
Chromosome 8 genomic sequence
unpublished
2 (bases 1 to 126753)
Genome Sequencing Center Jena.
Direct Submission
Submitted (04-APR-2000) Genome Analysis Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
On Jun 8, 2001 this sequence version replaced gi:8151669.
-----
Center: Institute of Molecular Biotechnology
```

```
Center code: IMB
Web site: http://genome.imb-jena.de/
Contact: gscf-submit@genome.imb-jena.de
-----
Project Information
Center project name: H3573M4
Center clone name: CFP-2373M4
-----
Summary Statistics
Sequencing vector: M13; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 121916 bases at least Q40
Consensus quality: 123958 bases at least Q30
Consensus quality: 125128 bases at least Q20
Quality coverage: 7.46 x in Q20 bases; sum-of-contigs
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality 10.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 3657: contig of 3657 bp in length.
* 3658 3757: gap of unknown length
* 3758 5367: contig of 50110 bp in length
* 5368 53967: gap of unknown length
* 53968 61529: contig of 7562 bp in length
* 61530 61629: gap of unknown length
* 61630 73303: contig of 11674 bp in length
* 73304 73403: gap of unknown length
* 73404 90798: contig of 17395 bp in length
* 90799 90898: gap of unknown length
* 90899 120417: contig of 29519 bp in length
* 120418 120517: gap of unknown length
* 120518 123861: contig of 3444 bp in length
* 123862 124061: gap of unknown length
* 124062 126753: contig of 2692 bp in length.
-----
FEATURES
source
1. 126753
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="8p12"
/clone="CTD-2373M4"
-----
BASE COUNT      37348 a      26850 c      26198 g      35657 t      700 others
ORIGIN
Query Match      35.5%; Score 124.2; DB 2; Length 126753;
Best Local Similarity 58.7%; Pred. No. 4.3e-09;
Matches 192; Conservative 29; Mismatches 98; Indels 8; Gaps 3;

QY 1 ATGAGCTCAGACTACTCTGACCTTACCCAGAGCCCTGAGTGGCCCGCTACCCCTC 60
    |||||
DB 27548 ATGAGCTCAGAGTGGCTGACCTTACCCAGAGCCCTGAGTGGCCCGCTACCCCTC 27607

QY 61 TGTACCCACGAGGAGTTAGCCGCTGCTTCTGCTGCTGCTTTCAGCTACCCGCTT 120
    |||||
DB 27608 TGTACCCACGAGGAGTTAGCCGCTGCTGCTGCTTTCAGCTACCCGCTT 27664

QY 121 CAGTCGATGCCAGAGGCGCTGTATACAT-GCCTAGAGNNNNACMCCTGCACNNNTGANA 179
    |||||
DB 27665 CAGTCGATGCCAGGAGTAAATGATACCAATCGGAGAGACTTACTGACTTCACACCACT 27724
```

Submitted: 123-NOV-2001) Whitehead Institute/MIT center for genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 1276577)
Birtten, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguski, L., Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeRellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fero, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Gange, S., Gord, S., Goyette, L., Graham, L., Grand-Pierre, N., Hages, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Mayor, J., Margusis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menais, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhthang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Riback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rossetti, M., Roy, A., Santos, R., Schauer, S., Schnupp, R.,

```

URES      Location/Qualifiers
source    1..127677
           /organism="Homo sapiens"
           /db_xref="taxon:9606"
           /chromosome="8"
           /map="8"
           /clone="CTD-2373N4"
repeat_region   /clone_lib="CTDI Human BAC"
               61..81
               /rpt_family="AT_rich"
repeat_region   135..285
               /rpt_family="FAM"
               290..318
               /rpl_family="(CA)n"
               578..612
               /rpl_family="(TG)n"
               complement(1567..1751)
               /rpl_family="MTIEI"
repeat_region   complement(1798..2049)
               /rpl_family="AlusX"
               2062..2099
               /rpl_family="(T)n"
               complement(2759..2844)
               /rpl_family="LIMC5"
repeat_region   complement(2845..3144)
               /rpl_family="Aluv"
repeat_region   complement(3165..3390)
               /rpl_family="LIMC5"
               complement(3577..3693)
repeat_region   /rpl_family="LIMC5"
               4068..4133
               /rpl_family="FLM_A"
               complement(4959..5211)
               /rpl_family="LMB2"
               5207..5245
               /rpl_family="LMB1"
               5277..5523
               /rpl_family="LMA8"
               5529..5660
               /rpl_family="FLM_A"
               5662..5696
               /rpl_family="Alub"
               5697..5725
               /rpl_family="(CAA)n"
               5726..5883
               /rpl_family="Alub"
               5902..5951
               /rpl_family="ALUB"
               5960..5982
               /rpl_family="AT_rich"
               5991..6138
               /rpl_family="LIMA10"
```

Query Match	35.58;	Score 124.2;	DB 5;	Length 1276777;
Best Local Similarity	58.78;	Pred. No. 4.3e-09;		

RESULT	15
Locus	AC011204
DEFINITION	AC011204 Homo sapiens, clone RP11-115P4, complete sequence.
ACCESSION	AC011204
VERSION	AC011204.6 GI:17048142
KEYWORDS	HTG;
PRI	25-NOV-2001

Organisms: *Homo sapiens*, *Enkariaula*, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

1 (bases 1 to 155926)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens, clone RP11-115P4
Unpublished
2 (bases 1 to 155926)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baldwin, J., Barna, N., Beckert, R., Bonnaville, L., Bonkigalter, B., Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeRellano, K., Dewar, K., Domingo, M., Donelan, L., Doyle, M., Farrell, P., FitzHugh, W., Forrest, C., Funke, K., Gage, D., Galagan, J., Gardua, S., Grant, G., Haas, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lehotzky, J., Lien, C., Locke, K., Macdonald, P., Margulis, N., McKean, P., McCork, A., McKernan, K., McLaughlin, J., Melnick, J., Morrow, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Seaver, P., Slange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W., Zimmer, A. and Zody, M.

Direct Submission
Submitted (03-OCT-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 155926)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,

Anderson, S., Barua, N., Bastien, V., Boguslavsky, L., Boukhalov, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, R., Chigwell, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeVellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardina, S., Grinde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Bajdos, R., Heard, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamal, A., Karatas, A., Keils, C., LaRoque, K., Lamatares, R., Landers, T., Lehoczy, J., Levine, R., Lin, G., Maclean, C., Macdonald, P., Major, J., Martin, N., Matthews, C., McCarthy, M., McLean, P., McKernan, K., McPheeters, R., Melidim, J.,

COMMENT

TITLE
JOURNAL

Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, R., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thoman, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, C., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (26-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 22, 2001 this sequence version replaced gi:16905307.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: 11917
Center clone name: 115_P4

FEATURES

source

```

1. 155926
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /clone_lib="RP11-115P4"
   /clone_id="RP11-115P4 Human Male BAC"
   /rpt_family="MIR"
   complement(1607. 1720)
   /rpt_family="12"
   complement(2259. 2563)
   /rpt_family="Alusx"
   complement(3701. 3835)
   /rpt_family="FLAM.C"
   complement(3856. 3928)
   /note="<30 qual SNGL region"
   /rpt_family="12"
   complement(4688. 4692)
   /note="<30 qual SNGL region"
   complement(4846. 4851)
   /note="<30 qual SNGL region"
   6859. 6901
   /note="single clone coverage"
   complement(6879. 7164)
   /rpt_family="Alusx"
   6902. 6906
   /note="<30 qual single clone coverage"
   6907. 6934
   /note="single clone coverage"
   7068. 7074
   /note="<30 qual SNGL region"
   7305. 7730
   /rpt_family="LTR56"
   complement(8498. 8797)
   /rpt_family="Alusq"
   complement(8800. 9116)
   /rpt_family="Alusx"
   complement(9325. 9630)
   /rpt_family="Alusq"
   complement(10345. 10620)
   /rpt_family="Alusp"
   10823. 11133

```

```

   /rpt_family="Alusp"
   complement(11677. 11921)
   /rpt_family="L1ME4A"
   complement(11922. 12217)
   /rpt_family="Aluoc"
   complement(12218. 12496)
   /rpt_family="L1ME4A"
   complement(12922. 13014)
   /rpt_family="MER104"
   complement(13180. 13350)
   /rpt_family="Aluoc"
   complement(13351. 13647)
   /rpt_family="Alusq"
   complement(13648. 13785)
   /rpt_family="Aluoc"
   14770. 15105
   /rpt_family="Aluub"
   15218. 15513
   /rpt_family="Aluub"
   15626. 15647
   /rpt_family="AT_rich"
   complement(15762. 16003)
   /rpt_family="Alusx"
   16046. 16090
   /rpt_family="(T)n"
   complement(16223. 16361)
   /rpt_family="12"
   complement(16412. 16630)
   /rpt_family="MIR"
   16987. 17285
   /rpt_family="Alusx"
   complement(17432. 17863)
   /rpt_family="L1MB8"
   complement(17929. 18231)
   /rpt_family="Alusp"
   complement(18362. 18669)
   /rpt_family="AlusC"
   complement(18737. 18858)
   /rpt_family="12"
   18883. 18928
   /rpt_family="(CA)n"
   complement(18942. 19244)
   /rpt_family="AlusC"
   19281. 19292
   /rpt_family="(TG)n"
   19293. 19318
   /rpt_family="(TG)n"
   complement(19460. 19774)
   /rpt_family="Aluoc"
   20119. 20428
   /rpt_family="Aluub"
   20441. 20751
   /rpt_family="Alusq"
   complement(21666. 21759)
   /rpt_family="MIR3"
   complement(21813. 22108)
   /rpt_family="Alusx"
   complement(22356. 22645)
   /rpt_family="Alusq"
   complement(22648. 22779)
   /rpt_family="Alusq/x"
   23151. 23222
   /rpt_family="tRNA-11e-ATW"
   23554. 24182
   /rpt_family="12"
   complement(25046. 25510)
   /rpt_family="Aluy"
   complement(25511. 25808)
   /rpt_family="Alusx"
   complement(25809. 25822)
   /rpt_family="Aluy"
   25860. 25863
   /rpt_family="AT_rich"

```

```

repeat_region complement(26434..26589)
repeat_region /rpl_family="LIMC/D"
repeat_region complement(26837..26984)
repeat_region /rpl_family="LIMC/D"
repeat_region complement(26985..27115)
repeat_region /rpl_family="FLAH_A"
repeat_region complement(27117..27298)
repeat_region /rpl_family="LIMC/D"
repeat_region 27307..27329
repeat_region /rpl_family="(TAA)n"

```

Query Match 35.58; Score 124.2; DB 9; Length 153526;
 Best Local Similarity 58.78; Pred. No. 4.6e-09;
 Matches 192; Conservative 29; Mismatches 98; Indels 8; Gaps 3;

```

QY 1 ATGAGCTCACAGTACCTGCTTACCCGAGTACCCCTGAGAGTGGGCGCCGCTACCCCTG 60
DB 153239 ATGAGCTCACAGTACCTGCTTACCCGAGTACCCCTGAGAGTGGGCGCCGCTACCCCTG 153298

QY 61 TGTACCCAGCGAGTTAGCGCCCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120
DB 153299 TGTACCCAGCGAGTTAGCGCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 153355

QY 121 CACTGCATGCCCGAGAGCGCCTGTATACAT GCCTAGGNNNNACAGCCTGAGNNNTGGANA 179
DB 153356 CACTGCATGCCCGAGAGTAAATTGATACCATCGGCGAGACTTTCACCTGACGACAGT 153415

QY 180 CTNCTNCAGCATGTGNATGATNCGTNCCTGGGAATGNCCTGNCAGATGTCAGCTACT 239
DB 153416 CCTTTCAAACTTGGTTCGCCAGAGCACAC---CATACAACTAACACTATACGAGCTAAC 153471

QY 240 GCAGACCAACNAGAGGAGCTTNGAGACAGAAATGATATNNAACNAGACTTATNCT 299
DB 153472 AGATCCACCCTCAAGAGATTAAATGATCCCTCAAAATCAAGCTGACACTCTTCAGGACT 153531

QY 300 TTGCAANCAANACTNTTCTTGGANA 326
DB 153532 GACGAGATGCTGTGACAGCAATGATA 153558

```

Search completed: May 9, 2003, 20:19:09
 Job time : 1525 secs

GenCore version 5.1.4.p3.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 9, 2003, 18:29:44 : Search time 219 Seconds
(without alignments)
3599.084 Million cell updates/sec

Title: US-09-880-711-328
Perfect score: 350
Sequence: 1 atgagccacagctactgca.....ttgaagatttcataaaaa (350)

Scoring table: IDENTITY_MUCDX
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125959159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50.4	25.8	845	21	AAC95479
2	65.2	28.6	632	24	ABN73285
3	65.2	28.6	632	24	ABN73375
4	59.4	27.0	3753	22	AAL04936
5	59.4	27.0	3753	23	ABL7830
6	58	26.6	617	23	ABV54245
7	58	26.6	16688	24	ABL32320
8	57.8	26.5	1782	22	AAF23414
9	57	25.3	8677	22	AAL36957

C 10	57	16.3	8680	22	AAL36956	Human musculoskele
C 11	56.6	16.2	1771	22	AAC88259	African Green Monk
C 12	56.6	16.2	165199	24	ABK83460	Human cDNA differe
C 13	56.4	16.1	820	24	ABN89955	Arabidopsis thaliana
C 14	56.4	16.1	831	21	AAC51574	Arabidopsis thaliana
C 15	56.4	16.1	843	21	AAC51564	Arabidopsis thaliana
C 16	56.4	16.1	1765	22	AAK82401	Human secreted pro
C 17	56.4	16.1	9060	12	AAO15132	Zucchini ACG synth
C 18	56.4	16.1	9060	19	AAV15702	Zucchini ACG synth
C 19	56.4	16.1	9060	22	AAD04542	Zucchini CP-ACG 1A
C 20	56.4	16.1	9060	22	AAF23619	Zucchini ACG synth
C 21	55.8	15.9	31952	22	AAK89370	Human digestive sy
C 22	55.6	15.9	3515	20	AAK60106	Arabidopsis thaliana
C 23	55.6	15.9	4874	20	AAK60103	Arabidopsis thaliana
C 24	55.4	15.8	2544	23	ABL21414	Prosopilla melanog
C 25	55.4	15.8	3454	23	ABLO9688	Prosopilla melanog
C 26	55.4	15.8	5759	23	ABL22540	Prosopilla melanog
C 27	55.2	15.8	12400	23	ABL07678	Prosopilla melanog
C 28	55.2	15.8	13001	22	AAK82979	Human immune/haema
C 29	55.2	15.8	32185	22	AAK16806	Human musculoskele
C 30	55	15.7	3409	23	AAK75766	DNA encoding novel
C 31	55	15.7	40267	24	AAK18958	Human cadherin-11k
C 32	55	15.7	40267	24	AAK88125	Human osteoblast d
C 33	54.8	15.7	396	22	AAK187801	Human polynucleoti
C 34	54.8	15.7	621	21	AAC51567	Arabidopsis thaliana
C 35	54.8	15.7	777	21	AAC13950	Arabidopsis thaliana
C 36	54.8	15.7	845	21	AAC35459	Arabidopsis thaliana
C 37	54.8	15.7	879	21	AAC52193	Arabidopsis thaliana
C 38	54.8	15.7	954	20	AAK30342	DNA encoding a hum
C 39	54.8	15.7	1138	24	AAK562431	cDNA sequence #218
C 40	54.8	15.7	1496	24	ABL89681	Human polynucleoti
C 41	54.8	15.7	1976	24	AAK62289	cDNA sequence #76
C 42	54.8	15.7	2035	19	AAV30292	Bacillus thuringie
C 43	54.8	15.7	2119	24	AAK31104	Human Transcription
C 44	54.8	15.7	2720	22	AAH18544	Human cDNA sequenc
C 45	54.8	15.7	85680	21	AAF22299	BAC containing rep

ALIGNMENTS

RESULT 1	
AAC95479	standard; cDNA: 845 BP.
XX	AAC95479:
XX	21-FEB-2001 (first entry)
XX	Human secreted protein gene 18 spq ID NO:28.
DE	Human: secreted protein; immunosuppressive; antiarthritic; vulnary;
KW	anti-rheumatic; antiproliferative; cytosolic; cardiac; vasotropic;
KW	cerebroprotective; neurotropic; antiproliferative; neoplasm;
KW	limbicide; ophthalmological; hyperproliferative disorder; neoplasm;
KW	autoimmune disease; rheumatoid arthritis; cardiovascular disorder;
KW	cerebrovascular disorder; cerebral ischemia; angiogenesis; vital;
KW	nervous system disorder; Alzheimer's disease; bacteria; infection;
KW	cell proliferation; skin aging; wound healing; chemotaxis;
KW	food additive; ss.
XX	
OS	Homo sapiens.
XX	
XX	WO200058334-A1.
PD	05-OCT-2000.
XX	
XX	22-MAR-2000: 2000WO-US07507.
PF	
XX	
PR	26-MAR-1999: 99US-0126594.
PR	17-DEC-1999: 99US-0172408.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.

[illegible]

KW		Bovine;	Bos taurus; EST; expressed sequence tag; totipotence;
XX			
OS		Bos taurus.	
XX			
FN		M0200194550-A2:	
PD		13-DEC-2001.	
XX			
PJ		07-JUN-2001; 2001WO-USI8576.	
PF			
PR		07-JUN-2000; 2000US-209874P.	
PR		06-JUN-2001; 2001US-0876143.	
PA	(INF1 -)	INFIGEN INC.	
PI	Eilertsen KJ,	Felster-Genskow M, Childs L;	
DR	WPT: 2002-351289/38.		
XX			
PP	An expressed sequence tag (EST),	the expression of which, or its complementary sequence, in a cell identifies the cell as a developmentally competent or incompetent cell	
PS	Example 16, page 142-143;	584pp; English.	
CC	The present invention describes an expressed sequence tag (EST), where the EST is an isolated, enriched, or purified nucleic acid sequence representing all or part of a gene, the expression of which, or its complementary sequence, in a cell identifies the cell as a developmentally competent or incompetent cell. Molecules which induce totipotency in one or more cells. Molecules which induce developmental incompetence in a cell line are useful for preventing a full term pregnancy in an animal and inhibiting totipotency. The molecules are also useful for treating a disease in an animal by inducing development of one or more cells of the animal into a specific cell type. The present sequence represents a bovine EST which is given in the exemplification of the present invention.		
SQ	Sequence 632 BP; 191 A; 124 C; 152 G; 161 T; 4 other;		
Query Match	18.6%; Score 65.2; DB 24; Length 632;		
Best Local Similarity	48.2%; Pred. No. 0.85;		
Matches 105; Conservative 29; Mismatches 63; Indels 21; Gaps 1			
OY	AGSNNNACGNCCTGGCANNNTGCANACTGNACCATGTGATNGCGTCNCTGCCA 213		
DJ	297 AGGCAAAAGCCCCACACCCTGTTTGTTCCTTCCTCACCAAATAATTGTCTGGGA 238		
OY	ATANGCCTGNCAGATGCAACCTACTTGACAAGCCCANCNAAGGAGCTTNNGACAGAATG 273		
DJ	237 ATGTGTTTTGAACATATCAACCTACGTACAGACAGAGAGAGAGCTCCCATGTTGAATTT 178		
OY	274 TAATANAACACNACTACCTTA-----TNCCTTTGCAANAAGANA 312		
DJ	177 GCCTTAGCTGTTTTGTTGTGTTGTTGTTTTCCTTCCTCCCTTTTGCAAAA 118		
OY	313 ACTNTTCCTTGANAAANTTTGAGAGATTTCCAATAAAAA 350		
DJ	117 ACTATTCTTGTAACGTTTGAGAGATTTCAATAAAAA 80		
RESULT 3			
ABN73375/C			
ID . ABN73375 standard; CDNA; 632 BP.			
AC	ABN73375:-		
DT	03-JUL-2002 (first entry)		
DE	Bovine embryonic germ (EG) cell cDNA EST 000203a CONFIG 59.		
KW	Bovine; Bos taurus; EST; expressed sequence tag; totipotence;		

KW development: gene; ss.
XX
OS Bos taurus.
XX
PN WO200194550-A2.
XX
PD 13-DEC-2001.
XX
PF 07-JUN-2001: 2001WO-US18576.
XX
PR 07-JUN-2000: 2000US-209674P.
PR 06-JUN-2001: 2001US-0876143.
XX
PA (INF1-) INF1GEN INC.
XX
PI Ellertsen KJ, Pfister-Genskow M, Childs L;
XX
DR WPL: 2002-351289/38.
XX
PT An expressed sequence tag (EST), the expressor of which, or its
PT complementarily sequence, in a cell identifies the cell as a
PT developmentally competent or incompetent cell.
XX
PS Example 16: Page 160; 584pp; English.
XX
CC The present invention describes an expressed sequence tag (EST), where
CC the EST is an isolated, enriched, or purified nucleic acid sequence
CC representing all or part of a gene, the expression of which, or its
CC complementarily sequence, in a cell identifies the cell as a
CC developmentally competent or incompetent cell. Molecules which induce
CC developmental competence in a cell line are useful for inducing
CC totipotency in one or more cells. Molecules which induce developmental
CC incompetence in a cell line are useful for preventing a full term
CC pregnancy in an animal and inhibiting totipotency. The molecules are
CC also useful for treating a disease in an animal by inducing development
CC of one or more cells of the animal into a specific cell type. The
CC present sequence represents a bovine EST which is given in the
CC exemplification of the present invention.
XX
SO Sequence 632 BP; 191 A; 124 C; 152 G; 161 T; 4 other:

Query Match 18.6%; Score 65.2; DB 26; Length 632;
Best Local Similarity 48.2%; Pred. No. 0.85;
Matches 105; Conservative 29; Mismatches 63; Indels 21; Gaps 1;

QY 154 ACGNNNNACNCCCTGACNNTGGANACTGTCAGCAGCATGATGATNCCGTCNCGCA 213
DB 297 AGCGAAAGCCGACACACCTGTTTTCCTCTTCCTCCAGCAATAGATTTGCTCTGGGA 238
QY 214 ATGNGCCCTGNCACAGATCACTACTGACAGCCAGCAGCAGAGCTTTNNGACAGCAATG 273
DB 237 ATGTGTTGTAGATATCAACCTATGTCAGACGACGAGAGCTGCGATTTGGAATTT 178
QY 274 TATTANAAACNACACCTTA-----TNCCTTTTGCAGAAAGANA 312
DB 177 GCTGTACCTGTTTGTCTGTTGTTGTTTTCCTTCCTTCCTTCCTTTGGCAAA 118
QY 313 ACTNTCTTGGANAATNTTGGAGATTCATATAAA 350
DB 117 ACTATTCTTGCTAGCGTTTGAGACATTCATAAATAA 80

RESULT 4
AAL04936/C
ID AAL04936 standard; DNA: 3753 BP.
XX
AC AAL04936;
XX
PT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA seq ID NO: 7624.
XX
KW Human: reproductive system related antigen; reproductive system disorder;

KW cancer: gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001: 2001WO-US01339.
XX
PR 31-JAN-2000: 2000US-0179065.
PR 04-FEB-2000: 2000US-0180628.
PR 24-FEB-2000: 2000US-0184664.
PR 02-MAR-2000: 2000US-0186350.
PR 16-MAR-2000: 2000US-0189874.
PR 17-MAR-2000: 2000US-0190076.
PR 18-APR-2000: 2000US-0198123.
PR 19-MAY-2000: 2000US-0205515.
PR 07-JUN-2000: 2000US-0209467.
PR 28-JUN-2000: 2000US-0214886.
PR 30-JUN-2000: 2000US-0215135.
PR 07-JUL-2000: 2000US-0216647.
PR 07-JUL-2000: 2000US-0216880.
PR 11-JUL-2000: 2000US-0217487.
PR 14-JUL-2000: 2000US-0217496.
PR 26-JUL-2000: 2000US-0218290.
PR 26-JUL-2000: 2000US-0220963.
PR 26-JUL-2000: 2000US-0220964.
PR 14-AUG-2000: 2000US-0224518.
PR 14-AUG-2000: 2000US-0224519.
PR 14-AUG-2000: 2000US-0225213.
PR 14-AUG-2000: 2000US-0225214.
PR 14-AUG-2000: 2000US-0225266.
PR 14-AUG-2000: 2000US-0225267.
PR 14-AUG-2000: 2000US-0225268.
PR 14-AUG-2000: 2000US-0225270.
PR 14-AUG-2000: 2000US-0225447.
PR 14-AUG-2000: 2000US-0225757.
PR 14-AUG-2000: 2000US-0225758.
PR 14-AUG-2000: 2000US-0225759.
PR 18-AUG-2000: 2000US-0226379.
PR 22-AUG-2000: 2000US-0226681.
PR 22-AUG-2000: 2000US-0226868.
PR 22-AUG-2000: 2000US-0227182.
PR 23-AUG-2000: 2000US-0227009.
PR 30-AUG-2000: 2000US-0228924.
PR 01-SEP-2000: 2000US-0229287.
PR 01-SEP-2000: 2000US-0229343.
PR 01-SEP-2000: 2000US-0229344.
PR 01-SEP-2000: 2000US-0229345.
PR 05-SEP-2000: 2000US-0229505.
PR 05-SEP-2000: 2000US-0229513.
PR 06-SEP-2000: 2000US-0230437.
PR 06-SEP-2000: 2000US-0230438.
PR 08-SEP-2000: 2000US-0231242.
PR 08-SEP-2000: 2000US-0231243.
PR 08-SEP-2000: 2000US-0231244.
PR 08-SEP-2000: 2000US-0231413.
PR 08-SEP-2000: 2000US-0231414.
PR 08-SEP-2000: 2000US-0231414.
PR 08-SEP-2000: 2000US-0232080.
PR 12-SEP-2000: 2000US-0232081.
PR 12-SEP-2000: 2000US-0232397.
PR 14-SEP-2000: 2000US-0232397.
PR 14-SEP-2000: 2000US-0232398.
PR 14-SEP-2000: 2000US-0232399.
PR 14-SEP-2000: 2000US-0232400.
PR 14-SEP-2000: 2000US-0232401.
PR 14-SEP-2000: 2000US-0233063.
PR 14-SEP-2000: 2000US-0233064.
PR 14-SEP-2000: 2000US-0233065.
PR 21-SEP-2000: 2000US-0234223.
PR 21-SEP-2000: 2000US-0234274.
PR 25-SEP-2000: 2000US-0234957.

[illegible]

(C) (T) E T O C O N

PA (EPiG-) EPIGENOMICS AG.

PA (EPiG-) EPIGENOMICS AG.

XX
P1 Olek A. Plepenbrock C, Berlin K:
XX
DR WPI: 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
XX
PS Claim 1: SEQ ID NO 293: 32pp + Sequence Listing: German.
XX
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anemia, cancer, acute myeloid
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/alicerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
XX
S0 Sequence 16688 BP: 4441 A: 200 C: 3799 G: 8248 T: 0 other:

Query Match 16.6%, Score 58; DB 24: Length 16688;
Best Local Similarity 41.2%; Pred. No. 8.6e-02;
Matches 80: Conservative 29; Mismatches 85; Indels 0; Gaps 0;

OY 157 NNNNACGCGACGNNNNGANACTGCTGACACATGATGCGCTGCTGGCAATG 216
DB 882 ATCTTAATAATACATCAATAATAATAATAACACTCTTAATAATAATAATAATACTTCTACTG 823
OY 217 NCCCTGACAGATACACCTACTGCGAGACGACGACGAGCTTNNAGACAGAAATGTAT 276
DB 822 AACTATACACTTAATCAACTACGCTTAATAAATCTAATAAATCTGATCTTAATAAAAAA 763
OY 277 TANNACGACTACTCTTATTCCTTTGCAANGANAACTGTTTGGANAATTTTGACA 336
DB 762 TATTAAAAAACTAAATCATCTCTATTAATAAAAAAAAGCTCTTAATAAAAAA 703
OY 337 GATTTCATTAATAA 350
DB 702 CATATCTTATTAATA 689

RESULT 8
AAF23414
ID AAF23414 standard: DNA: 1782 BP.
XX
AC AAF23414:
XX
DT 20-MAR-2001 (first entry)
DE Human SEC5 DNA sequence clone 3581580.0.30 SEQ ID 5.
XX
KW SEC5; secreted protein; cancer; angiogenesis; wound healing;
KW immune disorder; neurodegenerative disease; allergic reaction;
KW respiratory problem; organ transplantation; contraceptive; human;
KW chromosome 8; proliferative disorder; ds.
XX
OS Homo sapiens.
XX
PN WO200070046 A2.
XX
PD 23-NOV-2000.
XX
PE 12-MAY-2000: 2000WO-US13251.
XX
XX 14-MAY-1999: 99US-0134315.
PR 12-JAN-2000: 2000US-0175744.
PR 10-MAR-2000: 2000US-0188274.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkels RA, Fernandes E, Holdog F;

XX
DR WPI: 2001-025020/03.
XX
DR P-PSDB: AAB49553.
XX
XX New SPGX polypeptides and nucleic acids useful for treating or
PT preventing cancer, other disorders related to angiogenesis,
PT neurodegenerative diseases, autoimmune disorders and allergic reactions
PT -
XX
XX
PS Claim 9: Page 20-21: 132pp; English.
XX
XX
CC Polynucleotide sequences AAF23410 - AAF23419 encode secreted SPGX
CC proteins AAB49549 - AAB49558. Sequences AAF23420 - AAF23450 represent
CC primers and probes used in the isolation and characterization of the SECX
CC DNA sequences of the invention. The new polypeptides and nucleic acids
CC can be used in screening assays, detection assays, preventive or
CC predictive medicine, therapeutic and prophylactic treatment, and
CC pharmacogenomics. Specifically, the SECX polypeptides and nucleic acids
CC are useful for treating cancer; other disorders related to angiogenesis
CC e.g. abnormal wound healing, psoriasis; neurodegenerative diseases;
CC immune disorders; liver cirrhosis; benign tumours; fibrocystic conditions
CC and tissue hypertrophy (e.g. benign prostatic hypertrophy); allergic
CC reactions and conditions such as asthma and other respiratory problems;
CC as well as in treating or preventing diseases associated with organ
CC transplantation, arteriosclerosis-associated diseases or disorders. The
CC polypeptides can also be used for bone, cartilage, tendon, ligament
CC and/or tissue growth or regeneration, wound healing, tissue repair and
CC replacement, gut protection or regeneration, as a contraceptive, to
CC inhibit thromboses, infections caused by bacteria, virus, fungi and other
CC parasites, and as a vaccine. SECX antibodies may be used to isolate or
CC detect SECX proteins, monitor protein level in tissue as part of a
CC clinical testing procedure, treat proliferative disorders including
CC tumours and benign hyperplasias.
XX
XX
S0 Sequence 1782 BP: 521 A: 416 C: 397 G: 448 T: 0 other:

Query Match 16.5%; Score 57.8; DB 22: Length 1782;
Best Local Similarity 40.5%; Pred. No. 62;
Matches 83: Conservative 30; Mismatches 92; Indels 0; Gaps 0;

OY 139 NCTGTATATATGCGTAGGNNNCGACGCTGACGNNNNGANACTGCTGACGACATGATG 198
DB 185 AGTGATGCTGAATAATAATATTCATTGAGTTGATTTGTTGAGTATGCGCCAGGAATGCG 244
OY 199 ATNCGCTGCTGGGCAATGCGCTGACGAGATCAACTATCTGACACGACGACGACGACG 258
DB 245 ATGCTCAATCTGGAAGAGTAGGATGCCCATCAAGAGGAAAAACGACGACGATGACG 304
OY 259 TTNNCGAGACGAAATGATATTANNAACGACCTATATGCTTTGCAAAAGGAACTGTT 318
DB 305 CTACATGACGAGAAAGCTACGACAGCCCAATTTCAATATGAGTTGTAATTTGCAATGATGT 364
OY 319 CTTTGANAATTTTGGAGAGATTGCA 343
DB 365 CTTTAAATAATATGTCGCAAACTTGA 389

RESULT 9
AAL36957/C
ID AAL36957 standard: DNA: 8677 BP.
XX
AC AAL36957:
XX
DE 08-JAN-2002 (first entry)
XX
XX Human musculoskeletal system related polynucleotide SEQ ID NO 3322.
XX
XX
KW Cytostatic; immunosuppressive; neurologic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;
KW veterinary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ds.

XX	Rosen CA, Barash SC, Ruben SM;
PI	WPL: 2001-451937/48.
DR	Isolated polypeptide for treating, preventing and/or prognosing
XX	disorders related to the musculoskeletal system including
PT	musculoskeletal cancers and also for testing and detection e.g.
PT	diagnosis -
XX	
PS	Example 2: SEQ ID NO 3321; 781pp + Sequence Listing; English.
CC	The invention relates to novel genes (AAL34669-AAI37666) and proteins
CC	(ABB03087-ABB04109) associated with the musculoskeletal system useful
CC	for preventing, treating or ameliorating medical conditions e.g. by
CC	protein or gene therapy. The genes are isolated from a range of human
CC	tissues disclosed in the specification. The nucleic acids, proteins,
CC	antibodies and (anti)agonists are useful in the diagnosis, treatment
CC	and prevention of: (a) cancer, e.g. breast and ovarian cancer and
CC	other cancers of the adrenal gland, bone, bone marrow, breast,
CC	gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC	disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC	anemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC	multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
CC	(c) cardiovascular disorders such as myocardial ischemias; (d) wound
CC	healing; (e) neurological diseases such as cerebral anoxia and epilepsy;
CC	and (f) infectious diseases such as viral, bacterial, fungal and
CC	parasitic infections.
CC	Note: The sequence data for this patent did not form part of the
CC	printed specification, but was obtained in electronic format directly
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences.
SX	
SQ	Sequence 8680 BP; 2010 A; 2541 C; 2538 G; 1591 T; 0 other;
	Query Match 16.3% Score 57; DB 22; Length 8680;
	Best Local Similarity 40.2% Pred. No. 5.8e+02;
	Matches 84; Conservative 30; Mismatches 95; Indels 0; Gaps
OY	121 CACTGCATGCCACGAGGGCCTGTATCATCGCTAGGNNNCCNCCTGACNNNTGAAC 180
DY	: : : : :
DY	4979 CACTCAGGCCCTCACAGAGCTGTGTACTTCAAGCCCAGTCAGTAGACATTCTGGGCTC 4920
OY	181 TTCTCNACGACATCATGATGAGTGCGTCTGTGGAAATGNGCCTGCNCACAGATCAACTACTG 240
DY	4919 ACCGAGAGGGGCGCTGTGCTCCAGGAGCGGTGTGACCTGCAGAACAGGCATCTCTCC 4860
OY	241 CAGGCCACNMAAGGAGCTTTNAGACAGCATGTATTANNAACCACATCACTTAACCTT 300
DY	4859 AGGACCCCAAAGGGGAGCTGTAGGAGCCAGATGTCTGTACCGTCCGTCTGTCTGGAGC 4800
OY	301 TTGCAANGAANAACCTTTCTTTGANAAAT 329
DY	: :
DY	4799 TGACTGCTGCCCCCTTGCTTGATGTG 4771
RESULT 11	
AAC8299/c	
ID AAC8299 standard; DNA: 7171 BP.	
XX AAC8299;	
XX	
DT 02-MAR-2001 (first entry)	
DE African Green Monkey simian hepatitis A viral isolate AGM-27.	
XX African green monkey; hepatitis A; vaccine; attenuated strain;	
KW chimeric virus; ds.	
OS Simian hepatitis A virus.	
PN US6146643-A.	
DD 14-NOV-2000.	
XX	

DB 73402 GATCAATGACA 73392

RESULT 13

ID ABN98965 standard; DNA; 820 BP.

AC ABN98965;

XX

DT 01-AUG-2002 (first entry)

XX

DE Arabidopsis thaliana, expressed polynucleotide SEQ ID NO 733.

XX

KW Arabidopsis thaliana; plant; insecticide; fungicide; transgenic; stress; disease; crop; thale cress; tolerance factor; insect; pathogen; nutrition; ds.

XX

OS Arabidopsis thaliana.

XX

PN US2002023281-A1.

XX

PD 21-FEB-2002.

XX

PF 26-JAN-2001; 2001US-0770445.

XX

PR 27-JAN-2000; 2000US-178472P.

XX

PA (GORL/) GORLACH J.

PA (ANY/) AN Y.

PA (HAM/) HAMILTON C M.

PA (PRIC/) PRICE J L.

PA (RAIN/) RAINES T M.

PA (YUY/) YU Y.

PA (RAME/) RAMEKA J G.

PA (PAGE/) PAGE A.

PA (MATH/) MATHEN A V.

PA (LEDF/) LEDFORD B L.

PA (WOES/) WOESSNER J P.

PA (HAAS/) HAAS W D.

PA (GARC/) GARCIA C A.

PA (KRIC/) KRICKER M.

PA (SLAT/) SLATER T.

PA (DAVI/) DAVIS K R.

PA (ALE/) ALLEN K.

PA (HOFF/) HOFFMAN N.

PA (HURB/) HURBAN P.

XX

PI Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y, Rameka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD, Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N, Hurban P;

PI WPI: 2002-400781/43.

XX

PT New Arabidopsis thaliana nucleic acid for identifying homologous genes, producing compositions that modulate the expression or function of its encoded protein, and mapping functional regions of protein

XX

PS Claim 1: SEQ ID NO 733; 49pp + Sequence Listing; English.

XX

CC The invention relates to an Arabidopsis thaliana nucleic acid (I) comprising a sequence capable of hybridizing under stringent conditions to a sequence selected from any one of 999 sequences (ABN98233-ABN99231), given in the specification or its fragment. A polypeptide (II) encoded by (I), a transgenic plant (III) comprising an exogenous nucleic acid for a genetically modified cell (IV) comprising an exogenous nucleic acid, is useful for screening a candidate agent for its biological effect. (I) is useful in identifying homologous or related genes, in producing compositions that modulate the expression or function of its encoded protein, mapping functional regions of the protein and in studying associated physiological pathways. (I) is also useful for the genetic manipulation of cells, particularly plant cells. (I) is also useful in screening assays of various plant strains to determine the strains that

CC are best capable of withstanding a particular disease or environmental stress. (II) and (III) are useful for screening of biologically active agents, e.g. fungicides, insecticides, etc., for elucidating biochemical pathways. The screened agents are useful in improved methods of treating crops to prevent or treat disease. (II) are also useful in screening programs to identify agents that mimic or enhance the action of tolerance factors. Such agents are useful in improved methods of treating crops to enhance their tolerance to environmental stress. (I) is also useful for enhancing or inhibiting production of a biosynthetic product in a plant. (III) is useful for identifying other mediators that may induce expression of proteins of interest, for establishing the extent to which any specific insect and/or pathogen is responsible for damage to a particular plant, for identifying other mediators that enhance or induce tolerance to environmental stress, for identifying factors involved in biosynthetic pathways of nutritional, commercial, or medicinal value and for identifying productions of nutritional, commercial or medicinal value. (IV) is useful in the study of genetic function and regulation, for alteration of the cellular metabolism and for screening compounds that may affect the biological function of the gene or gene products. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?docID=999099770445.

XX

Seq Sequence 820 BP; 230 A; 185 C; 193 G; 212 T; 0 other;

Query Match 16.1%; Score 56.4; DB 24; Length 820;

Best Local Similarity 39.7%; Pred. No. 43;

Matches 96; Conservative 30; Mismatches 116; Indels 0; Gaps 0;

QY 88 TTCCTCCTCCCTCCGCTTACCTACCCCGCTTACCTGCATGCCAGAGCNCCTATAC 147

DB 189 TCCCTGCTCCACTCTGTGACCTCCGCCCTGTGTGATGTGATGACTCTTAAGTACA 248

QY 148 ATGCCTAGGNNNNCAACNCTGACNNNTGGANACTGNCAGACATGATGATNCCGTG 207

DB 249 ACATGAAGGTGACGAACCGGTAAAGGATTGACTTGAAGAGCTCAAGCTGTGATACC 308

QY 208 CTGGGATGNGCCTGNCACAGATCAACCTACCTGACGAGCAGCANNAGAGGCTTNGAGAC 267

DB 309 CAAGAAGTGTGGCCGCTCAATGTTGATGCTGTGACCATGTGCGAAGACCATCTT 368

QY 268 AGAATGTATTANNAACNNACTACCTTATNCTTTTGGCAANGANNACTNTCTTGANNA 327

DB 369 TGGAGGCTTCAACAATGTCAGAGGCTGAANAACCTACAGAAGCAAGTATGATTT 428

QY 328 TN 329

DB 429 TC 430

XX

RESULT 14

ID AAC51574

XX

AC AAC51574 standard; DNA; 831 BP.

XX

DT 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 69015.

XX

KW Hybridisation assay; genetic mapping; gene expression control; protein identification; signal transduction pathway; metabolic pathway; promoter; termination sequence; ss.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.

PR	05-MAR-1999:	990S-0123180.
PR	09-MAR-1999:	990S-0123548.
PR	23-MAR-1999:	990S-0125788.
PR	25-MAR-1999:	990S-0126264.
PR	29-MAR-1999:	990S-0126785.
PR	01-APR-1999:	990S-0127462.
PR	06-APR-1999:	990S-0128234.
PR	08-APR-1999:	990S-0128717.
PR	16-APR-1999:	990S-0130077.
PR	19-APR-1999:	990S-0130449.
PR	21-APR-1999:	990S-0130510.
PR	23-APR-1999:	990S-0130891.
PR	28-APR-1999:	990S-0131449.
PR	30-APR-1999:	990S-0132048.
PR	30-APR-1999:	990S-0132407.
PR	04-MAY-1999:	990S-0132484.
PR	05-MAY-1999:	990S-0132485.
PR	06-MAY-1999:	990S-0132486.
PR	06-MAY-1999:	990S-0132487.
PR	07-MAY-1999:	990S-0132866.
PR	11-MAY-1999:	990S-0134256.
PR	14-MAY-1999:	990S-0134218.
PR	14-MAY-1999:	990S-0134221.
PR	14-MAY-1999:	990S-0134370.
PR	18-MAY-1999:	990S-0134768.
PR	19-MAY-1999:	990S-0134941.
PR	20-MAY-1999:	990S-0135121.
PR	21-MAY-1999:	990S-0135353.
PR	24-MAY-1999:	990S-0135629.
PR	25-MAY-1999:	990S-0136021.
PR	27-MAY-1999:	990S-0136392.
PR	28-MAY-1999:	990S-0136782.
PR	01-JUN-1999:	990S-0137222.
PR	03-JUN-1999:	990S-0137528.
PR	04-JUN-1999:	990S-0137502.
PR	07-JUN-1999:	990S-0137724.
PR	08-JUN-1999:	990S-0138094.
PR	10-JUN-1999:	990S-0138540.
PR	10-JUN-1999:	990S-0138847.
PR	14-JUN-1999:	990S-0139111.
PR	16-JUN-1999:	990S-0139452.
PR	17-JUN-1999:	990S-0139453.
PR	18-JUN-1999:	990S-0139492.
PR	18-JUN-1999:	990S-0139455.
PR	18-JUN-1999:	990S-0139456.
PR	18-JUN-1999:	990S-0139461.
PR	18-JUN-1999:	990S-0139462.
PR	18-JUN-1999:	990S-0139463.
PR	18-JUN-1999:	990S-0139750.
PR	18-JUN-1999:	990S-0139763.
PR	21-JUN-1999:	990S-0139817.
PR	22-JUN-1999:	990S-0139899.
PR	23-JUN-1999:	990S-0140353.
PR	23-JUN-1999:	990S-0140354.
PR	24-JUN-1999:	990S-0140695.
PR	28-JUN-1999:	990S-0140823.
PR	29-JUN-1999:	990S-0140991.
PR	30-JUN-1999:	990S-0141287.
PR	01-JUL-1999:	990S-0141842.
PR	02-JUL-1999:	990S-0142154.
PR	06-JUL-1999:	990S-0142390.
PR	08-JUL-1999:	990S-0142800.
PR	09-JUL-1999:	990S-0142920.
PR	12-JUL-1999:	990S-0143297.
PR	13-JUL-1999:	990S-0143542.
PR	14-JUL-1999:	990S-0143624.
PR	15-JUL-1999:	990S-0144085.
PR	16-JUL-1999:	990S-0144086.
PR	19-JUL-1999:	990S-0144325.
PR	19-JUL-1999:	990S-0144331.
PR	19-JUL-1999:	990S-0144332.
PR	19-JUL-1999:	990S-0144333.
PR	19-JUL-1999:	990S-0144334.
PR	19-JUL-1999:	990S-0144335.
PR	20-JUL-1999:	990S-0144352.
PR	20-JUL-1999:	990S-0144632.
PR	20-JUL-1999:	990S-0144684.
PR	21-JUL-1999:	990S-0144841.
PR	21-JUL-1999:	990S-0145086.
PR	21-JUL-1999:	990S-0145088.
PR	22-JUL-1999:	990S-0145085.
PR	22-JUL-1999:	990S-0145087.
PR	22-JUL-1999:	990S-0145089.
PR	22-JUL-1999:	990S-0145192.
PR	23-JUL-1999:	990S-0145145.
PR	23-JUL-1999:	990S-0145218.
PR	23-JUL-1999:	990S-0145224.
PR	26-JUL-1999:	990S-0145276.
PR	27-JUL-1999:	990S-0145913.
PR	27-JUL-1999:	990S-0145918.
PR	27-JUL-1999:	990S-0145919.
PR	28-JUL-1999:	990S-0145951.
PR	02-AUG-1999:	990S-0146388.
PR	02-AUG-1999:	990S-0146388.
PR	02-AUG-1999:	990S-0146389.
PR	02-AUG-1999:	990S-0147038.
PR	04-AUG-1999:	990S-0147204.
PR	05-AUG-1999:	990S-0147302.
PR	05-AUG-1999:	990S-0147192.
PR	05-AUG-1999:	9

OS	Arbidolopsis thaliana.	
XX		
PN	EP1033405-A2.	
XX		
PD	06-SEP-2000.	
XX		
PF	25-FEB-2000; 2000EP-03014539	
XX		
PR	25-FEB-1999; 990US-01281255	
PR	05-MAR-1999; 990US-01231810	
PR	09-MAR-1999; 990US-01334860	
PR	23-MAR-1999; 990US-01257878	
PR	23-MAR-1999; 990US-01262524	
PR	29-MAR-1999; 990US-01267055	
PR	01-APR-1999; 990US-01274645	
PR	01-APR-1999; 990US-01274642	
PR	06-APR-1999; 990US-01282324	
PR	08-APR-1999; 990US-01287174	
PR	16-APR-1999; 990US-01298485	
PR	19-APR-1999; 990US-01300749	
PR	21-APR-1999; 990US-01300449	
PR	23-APR-1999; 990US-01305100	
PR	23-APR-1999; 990US-01305911	
PR	28-APR-1999; 990US-01311445	
PR	30-APR-1999; 990US-01312048	
PR	30-APR-1999; 990US-01324077	
PR	04-MAY-1999; 990US-01324845	
PR	05-MAY-1999; 990US-01324845	
PR	06-MAY-1999; 990US-01324846	
PR	06-MAY-1999; 990US-01324846	
PR	07-MAY-1999; 990US-01324847	
PR	11-MAY-1999; 990US-01328653	
PR	11-MAY-1999; 990US-01342265	
PR	14-MAY-1999; 990US-01342218	
PR	14-MAY-1999; 990US-01342919	

PR	25-FEB-2000;	2000EP-03014439
PR	25-FEB-1999;	9905-01218255
PR	05-MAR-1999;	9905-01231800
PR	09-MAR-1999;	9905-01253548
PR	23-MAR-1999;	9905-01257387
PR	26-MAR-1999;	9905-01267624
PR	29-MAR-1999;	9905-01267655
PR	01-APR-1999;	9905-01274672
PR	06-APR-1999;	9905-01282424
PR	08-APR-1999;	9905-01287124
PR	16-APR-1999;	9905-01298455
PR	19-APR-1999;	9905-01300749
PR	21-APR-1999;	9905-01300449
PR	23-APR-1999;	9905-01305100
PR	28-APR-1999;	9905-01308921
PR	28-APR-1999;	9905-01311449
PR	30-APR-1999;	9905-01320485
PR	30-APR-1999;	9905-01324077
PR	04-MAY-1999;	9905-01324854
PR	06-MAY-1999;	9905-01324855
PR	06-MAY-1999;	9905-01324865
PR	07-MAY-1999;	9905-01324847
PR	07-MAY-1999;	9905-01326853
PR	11-MAY-1999;	9905-01342626
PR	11-MAY-1999;	9905-01343218
PR	14-MAY-1999;	9905-01342119
PR	14-MAY-1999;	9905-01343221
PR	14-MAY-1999;	9905-01343731
PR	18-MAY-1999;	9905-01347461
PR	19-MAY-1999;	9905-01349411

OS	Arbidopsis thaliana.
PN	EP1033405-A2.
XX	
PF	25-FEB-2000; 2000EP-0301439
XX	
XX	
PR	25-FEB-1999; 9905-0121825
PR	05-MAR-1999; 9905-0123180
PR	09-MAR-1999; 9905-0123548
PR	23-MAR-1999; 9905-0125788
PR	25-MAR-1999; 9905-0126254
PR	29-MAR-1999; 9905-0126785
PR	01-APR-1999; 9905-0127462
PR	06-APR-1999; 9905-0128242
PR	08-APR-1999; 9905-0128845
PR	15-APR-1999; 9905-0129874
PR	19-APR-1999; 9905-0130077
PR	21-APR-1999; 9905-0130549
PR	23-APR-1999; 9905-0130510
PR	28-APR-1999; 9905-0130891
PR	30-APR-1999; 9905-0131149
PR	30-APR-1999; 9905-0132048
PR	04-MAY-1999; 9905-0132407
PR	05-MAY-1999; 9905-0132464
PR	06-MAY-1999; 9905-0132465
PR	06-MAY-1999; 9905-0132466
PR	07-MAY-1999; 9905-0132467
PR	11-MAY-1999; 9905-0132483
PR	14-MAY-1999; 9905-0134218
PR	14-MAY-1999; 9905-0134219
PR	14-MAY-1999; 9905-0134221
PR	14-MAY-1999; 9905-0134370
PR	18-MAY-1999; 9905-0134376
PR	19-MAY-1999; 9905-0134941
PR	20-MAY-1999; 9905-0135124
PR	21-MAY-1999; 9905-0135353
PR	25-MAY-1999; 9905-0136021
PR	25-MAY-1999; 9905-0136029
PR	27-MAY-1999; 9905-0136592
PR	28-MAY-1999; 9905-0136782
PR	01-JUN-1999; 9905-0137222
PR	03-JUN-1999; 9905-0137528
PR	04-JUN-1999; 9905-0137580
PR	07-JUN-1999; 9905-0137724
PR	08-JUN-1999; 9905-0138094
PR	10-JUN-1999; 9905-0138540
PR	10-JUN-1999; 9905-0138847
PR	14-JUN-1999; 9905-0139119
PR	16-JUN-1999; 9905-0139452
PR	17-JUN-1999; 9905-0139453
PR	18-JUN-1999; 9905-0139492
PR	18-JUN-1999; 9905-0139494
PR	18-JUN-1999; 9905-0139455

OS	Arbidopsis thaliana.	
XX	EpI033405-A2.	
PN		
XX		
XX	06-SEP-2000.	
PD		
XX		
PF	25-FEB-2000.	2000EP-030149
XX		
PR	25-FEB-1999.	9905-0121825
PR	05-MAR-1999.	9905-0123180
PR	04-MAR-1999.	9905-0123548
PR	23-MAR-1999.	9905-0125578
PR	25-MAR-1999.	9905-0126264
PR	29-MAR-1999.	9905-0126785
PR	01-APR-1999.	9905-0127462
PR	06-APR-1999.	9905-0128234
PR	08-APR-1999.	9905-0128714
PR	16-APR-1999.	9905-0129845
PR	19-APR-1999.	9905-0130077
PR	21-APR-1999.	9905-0130049
PR	23-APR-1999.	9905-0130510
PR	23-APR-1999.	9905-0130891
PR	26-APR-1999.	9905-0131449
PR	30-APR-1999.	9905-0132048
PR	03-MAY-1999.	9905-0132407
PR	04-MAY-1999.	9905-0132444
PR	05-MAY-1999.	9905-0132485
PR	06-MAY-1999.	9905-0132486
PR	07-MAY-1999.	9905-0132487
PR	11-MAY-1999.	9905-0132863
PR	14-MAY-1999.	9905-0134256
PR	14-MAY-1999.	9905-0134218
PR	14-MAY-1999.	9905-0134219
PR	14-MAY-1999.	9905-0134321
PR	14-MAY-1999.	9905-0134370
PR	18-MAY-1999.	9905-0134768
PR	19-MAY-1999.	9905-0135141
PR	20-MAY-1999.	9905-0135194
PR	21-MAY-1999.	9905-0135353
PR	24-MAY-1999.	9905-0135629
PR	27-MAY-1999.	9905-0136021
PR	28-MAY-1999.	9905-0136782
PR	01-JUN-1999.	9905-0137222
PR	03-JUN-1999.	9905-0137528
PR	04-JUN-1999.	9905-0137502
PR	07-JUN-1999.	9905-0137724
PR	08-JUN-1999.	9905-0138094
PR	10-JUN-1999.	9905-0138540
PR	10-JUN-1999.	9905-0138847
PR	14-JUN-1999.	9905-0139112
PR	16-JUN-1999.	9905-0139452
PR	17-JUN-1999.	9905-0139463
PR	18-JUN-1999.	9905-0139454
PR	18-JUN-1999.	9905-0139456
PR	18-JUN-1999.	9905-0139457
PR	18-JUN-1999.	9905-0139458
PR	18-JUN-1999.	9905-0139459
PR	18-JUN-1999.	9905-0139460
PR	18-JUN-1999.	9905-0139461
PR	18-JUN-1999.	9905-0139462
PR	18-JUN-1999.	9905-0139463
PR	18-JUN-1999.	9905-0139464
PR	18-JUN-1999.	9905-0139465
PR	22-JUN-1999.	9905-0139617
PR	23-JUN-1999.	9905-0140353
PR	23-JUN-1999.	9905-0140354
PR	24-JUN-1999.	9905-0140695
PR	28-JUN-1999.	9905-0140823
20-JUN-1999.		

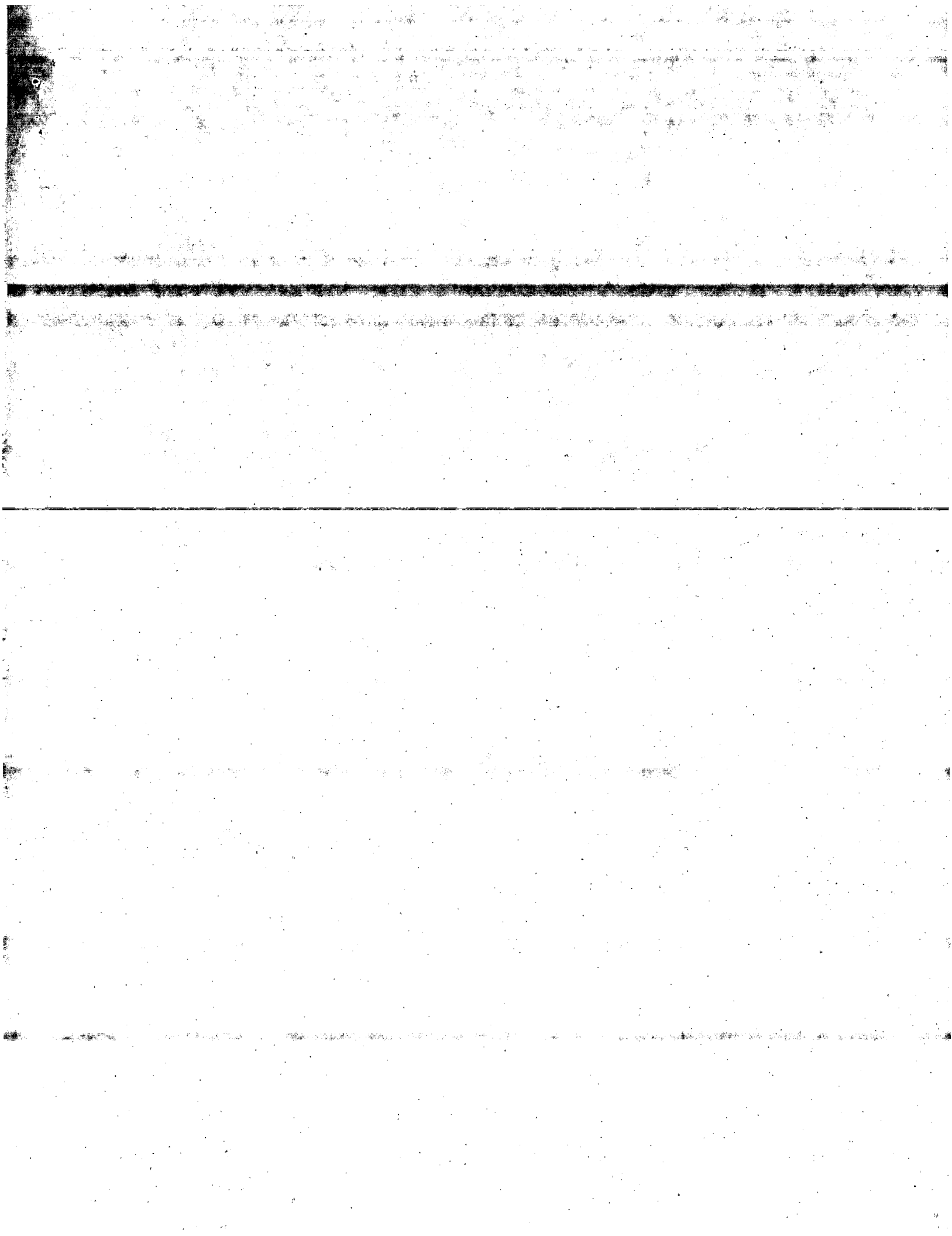
PR 30-JUN-1999: 990S-0141287.
 PR 01-JUL-1999: 990S-0141842.
 PR 01-JUL-1999: 990S-0142154.
 PR 02-JUL-1999: 990S-0142055.
 PR 06-JUL-1999: 990S-0142350.
 PR 08-JUL-1999: 990S-0142803.
 PR 09-JUL-1999: 990S-0142920.
 PR 12-JUL-1999: 990S-0142977.
 PR 13-JUL-1999: 990S-0143542.
 PR 14-JUL-1999: 990S-0143624.
 PR 15-JUL-1999: 990S-0144005.
 PR 16-JUL-1999: 990S-0144085.
 PR 16-JUL-1999: 990S-0144086.
 PR 15-JUL-1999: 990S-0144325.
 PR 15-JUL-1999: 990S-0144331.
 PR 19-JUL-1999: 990S-0144332.
 PR 15-JUL-1999: 990S-0144333.
 PR 19-JUL-1999: 990S-0144334.
 PR 19-JUL-1999: 990S-0144335.
 PR 20-JUL-1999: 990S-0144352.
 PR 20-JUL-1999: 990S-0144632.
 PR 20-JUL-1999: 990S-0144884.
 PR 21-JUL-1999: 990S-0144814.
 PR 21-JUL-1999: 990S-0145086.
 PR 21-JUL-1999: 990S-0145088.
 PR 22-JUL-1999: 990S-0145085.
 PR 22-JUL-1999: 990S-0145087.
 PR 22-JUL-1999: 990S-0145089.
 PR 22-JUL-1999: 990S-0145192.
 PR 23-JUL-1999: 990S-0145145.
 PR 23-JUL-1999: 990S-0145218.
 PR 23-JUL-1999: 990S-0145224.
 PR 26-JUL-1999: 990S-0145276.
 PR 27-JUL-1999: 990S-0145913.
 PR 27-JUL-1999: 990S-0145918.
 PR 28-JUL-1999: 990S-0145919.
 PR 28-JUL-1999: 990S-0145951.
 PR 02-AUG-1999: 990S-0146386.
 PR 02-AUG-1999: 990S-0146388.
 PR 03-AUG-1999: 990S-0146389.
 PR 03-AUG-1999: 990S-0147038.
 PR 04-AUG-1999: 990S-0147204.
 PR 04-AUG-1999: 990S-0147302.
 PR 05-AUG-1999: 990S-0147192.
 PR 05-AUG-1999: 990S-0147260.
 PR 06-AUG-1999: 990S-0147303.
 PR 06-AUG-1999: 990S-0147416.
 PR 09-AUG-1999: 990S-0147493.
 PR 09-AUG-1999: 990S-0147935.
 PR 10-AUG-1999: 990S-0148171.
 PR 11-AUG-1999: 990S-0148319.
 PR 12-AUG-1999: 990S-0148341.
 PR 13-AUG-1999: 990S-0148565.
 PR 13-AUG-1999: 990S-0148684.
 PR 16-AUG-1999: 990S-0149368.
 PR 17-AUG-1999: 990S-0149175.
 PR 18-AUG-1999: 990S-0149426.
 PR 20-AUG-1999: 990S-0149722.
 PR 20-AUG-1999: 990S-0149723.
 PR 20-AUG-1999: 990S-0149929.
 PR 23-AUG-1999: 990S-0149902.
 PR 23-AUG-1999: 990S-0149930.
 PR 25-AUG-1999: 990S-0150566.
 PR 26-AUG-1999: 990S-0150884.
 PR 27-AUG-1999: 990S-0151065.
 PR 27-AUG-1999: 990S-0151066.
 PR 27-AUG-1999: 990S-0151080.
 PR 30-AUG-1999: 990S-0151303.
 PR 31-AUG-1999: 990S-0151438.
 PR 01-SEP-1999: 990S-0151930.
 PR 07-SEP-1999: 990S-0152363.
 PR 10-SEP-1999: 990S-0153070.
 PR 13-SEP-1999: 990S-0153758.

PR 15-SEP-1999: 990S-0154018.
 PR 16-SEP-1999: 990S-0154039.
 PR 20-SEP-1999: 990S-0154779.
 PR 22-SEP-1999: 990S-0155133.
 PR 23-SEP-1999: 990S-0155186.
 PR 24-SEP-1999: 990S-0155659.
 PR 28-SEP-1999: 990S-0156458.
 PR 29-SEP-1999: 990S-0156596.
 PR 04-OCT-1999: 990S-0157117.
 PR 05-OCT-1999: 990S-0157753.
 PR 06-OCT-1999: 990S-0157865.
 PR 07-OCT-1999: 990S-0158029.
 PR 08-OCT-1999: 990S-0158232.
 PR 12-OCT-1999: 990S-0158365.
 PR 13-OCT-1999: 990S-0159293.
 PR 13-OCT-1999: 990S-0159294.
 PR 13-OCT-1999: 990S-0159295.
 PR 14-OCT-1999: 990S-0159329.
 PR 14-OCT-1999: 990S-0159329.
 PR 14-OCT-1999: 990S-0159330.
 PR 14-OCT-1999: 990S-0159331.
 PR 14-OCT-1999: 990S-0159637.
 PR 14-OCT-1999: 990S-0159638.
 PR 18-OCT-1999: 990S-0159584.
 PR 18-OCT-1999: 990S-0160741.
 PR 21-OCT-1999: 990S-0160757.
 PR 21-OCT-1999: 990S-0160768.
 PR 21-OCT-1999: 990S-0160770.
 PR 21-OCT-1999: 990S-0160814.
 PR 21-OCT-1999: 990S-0160815.
 PR 22-OCT-1999: 990S-0160980.
 PR 22-OCT-1999: 990S-0160981.
 PR 22-OCT-1999: 990S-0160989.
 PR 25-OCT-1999: 990S-0161404.
 PR 25-OCT-1999: 990S-0161405.
 PR 25-OCT-1999: 990S-0161406.
 PR 26-OCT-1999: 990S-0161359.
 PR 26-OCT-1999: 990S-0161360.
 PR 26-OCT-1999: 990S-0161361.
 PR 28-OCT-1999: 990S-0161920.
 PR 28-OCT-1999: 990S-0161592.
 PR 28-OCT-1999: 990S-0161593.
 PR 29-OCT-1999: 990S-0162142.

Query Match 16.1%: Score 56.4; DB 21; Length 843;
 Best Local Similarity 39.7%: Pred. No. 44;
 Matches 96; Conservative 30; Mismatches 116; Indels 0; Gaps 0;

QY 88 TWGCTGCTGCTGCGGCTTGACCTGACCGCGCTTCACTGCATGCGCAGAGGCTGTATAC 147
 DB 202 TCCCTGCTGCTGACCTGCTGACCTGCGGCTGTGTCATGCTGACCTTAACTACA 261
 QY 148 ATGCGCTAGGNNNNACNCGTGCAGNNNNYGGANACTTCTNCAGCATGATNGCTGCTN 207
 DB 262 ACATGAGGCTGACAGCCGGTGAAGGATTCAGCTTTCAGAGGCTCAAGGGCTGGATCC 321
 QY 208 CTGCGATATGCTGCTGACGATGACCTAGCTGACGACGANCNAAAGGACCTTNNCAAC 267
 DB 322 CAACGAGATTTGGCGCTACATATTTGCTATTCGTTTACCATCTGCGCAACAGCATTT 381
 QY 268 AGAATCTATTANNAACNNACTTATATNCCTTTTGCAANGAANGAATTTCTTTAANA 327
 DB 382 TGGAGGCTTTCAAACAAATGCTCAGAGGCTGAAGACCTGACAGCAAGTATGATTT 441
 QY 328 TTN 329
 DB 442 TC 443

Search completed: May 5, 2003, 15:54:51
 Job time : 295 secs



GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 9, 2003, 19:35:14 : Search time 1497 Seconds
(without alignments)
3786.520 Million cell updates/sec

Title: US-09-880-711-328

Perfect score: 350

Sequence: 1 atgagctcagctacgtcga.....ttgagagcttcacataaaa 350

Scoring table: IDENTITY_NUCDX

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_estbam:*
3: em_estlin:*
4: em_estmu:*
5: em_estcov:*
6: em_estcpl:*
7: em_estcro:*
8: em_estc:*
9: qb_est1:*
10: qb_est2:*
11: qb_est3:*
12: qb_est4:*
13: qb_est5:*
14: qb_est6:*
15: em_estlin:*
16: em_estom:*
17: qb_est:*
18: em_qss_hum:*
19: em_qss_inv:*
20: em_qss_pin:*
21: em_qss_vrl:*
22: em_qss_fun:*
23: em_qss_mam:*
24: em_qss_mus:*
25: em_qss_other:*
26: em_qss_pro:*
27: em_qss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	250.8	71.7	558	12	BF118340 uy93h12.x
C 2	249.2	71.2	442	12	BE948361 UI-M-BH3
C 3	249.2	71.2	787	12	BC069460 H3075H05-
C 4	249.2	71.2	1040	11	AK012265 Mus muscu
C 5	248.2	70.9	1040	11	AK012586 Mus muscu
C 6	244.4	69.8	420	5	AA960653 ub60c05.s

C 7	241.6	69.0	423	9	AU020489	AU020489
C 8	238.2	68.1	460	9	AU019051	AU019051
C 9	237.2	67.8	607	12	BE949377	BE949377 UI-M-BH3-
C 10	225.2	64.3	383	5	AA958210	AA958210 UI-R-CO-1
C 11	225.2	64.3	410	9	A1712872	A1712872 UI-R-AG1-
C 12	225.2	64.3	410	13	B1288680	B1288680 UI-R-DK0-
C 13	225.2	64.3	423	13	B1293426	B1293426 UI-R-DK0-
C 14	225.2	64.3	559	9	A1103164	A1103164 EST212453
C 15	225.2	64.3	584	5	A1231308	A1231308 EST227996
C 16	224.4	64.1	392	5	A1712611	A1712611 UI-R-AP1
C 17	223.6	63.9	462	10	AA434801	AA434801 UI-R-BUP
C 18	223.6	63.9	472	12	BF395018	BF395018 UI-R-CM0-
C 19	215.6	61.6	612	9	AA799579	AA799579 EST189076
C 20	207.6	59.3	355	5	A1704461	A1704461 UI-R-ACO-
C 21	201.2	57.5	412	5	A1339365	A1339365 CL15404.X
C 22	201.2	57.5	413	10	AA468894	AA468894 BC74403.X
C 23	201.2	57.5	425	12	BC163526	BC163526 602338550
C 24	201.2	57.5	462	12	BF939618	BF939618 7047108.X
C 25	201.2	57.5	489	9	A1347440	A1347440 q958a01.X
C 26	201.2	57.5	529	10	AA516503	AA516503 xp88a01.X
C 27	201.2	57.5	551	5	A1758331	A1758331 U06d11.X
C 28	201.2	57.5	553	9	A1458831	A1458831 U174609.X
C 29	201.2	57.5	560	14	BM682266	BM682266 UI-E-E10-
C 30	201.2	57.5	575	13	BM653438	BM653438 UI-E-C10-
C 31	201.2	57.5	645	9	A1472315	A1472315 U387c01.X
C 32	201.2	57.5	656	9	A1921608	A1921608 w027104.X
C 33	201.2	57.5	804	12	BE744630	BE744630 601577462
C 34	199.6	57.0	380	10	AA078977	AA078977 XB21002.X
C 35	199.6	57.0	593	14	B0021517	B0021517 UI-H-DH1-
C 36	199.6	57.0	623	9	A1715649	A1715649 A14604.X
C 37	197.4	55.4	348	9	A1713665	A1713665 UI-R-AG1-
C 38	197.2	55.3	460	13	B1402360	B1402360 M1-P-CPO-
C 39	196.4	55.1	303	9	AA667227	AA667227 vq89n07.F
C 40	196.4	55.1	723	9	A1749254	A1749254 A141603.X
C 41	196.4	55.1	726	10	AA001185	AA001185 wu25e11.X
C 42	193.2	55.2	429	9	AA045722	AA045722 zK81f10.F
C 43	193.2	55.2	753	13	B1831537	B1831537 603074463
C 44	193.2	55.2	957	14	B0931566	B0931566 AGENICOURT
C 45	191.2	54.6	294	9	AA866678	AA866678 nd07b11.F

ALIGNMENTS

RESULT 1
BF118340/c 558 bp mRNA linear EST 27-DEC-2000
LOCUS uy93h12.x1 NC1_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3667175 3
DEFINITION similar to YR:092517 092517 RRP-MS/TYPF 3, //, mRNA sequence.
ACCESSION BF118340
VERSION BF118340.1 GI:10587816
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerogastri; Muridae; Murinae; Mus.
REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS Tissue Procurement: Jolhar Hennighausen Ph.D., Robin Humphreys
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the T.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov/image/blast/resources.shtml>
COMMENT JOURNAL: uy93h12.y1
Other ESTs: uy93h12.y1
Contact: Robert Strausberg, Ph.D.
Email: graphs-tr@mail.nih.gov
Tissue Procurement: Jolhar Hennighausen Ph.D., Robin Humphreys
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the T.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov/image/blast/resources.shtml>
MG1:1427943

Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Alzawa, T., Hara, A., Fukunishi, Y., Konno, H., Akauchi, J., Fukuda, S.,
Alzawa, K., Iizawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamagata, I.,
Saito, T., Okazaki, Y., Gogjorian, T., Bono, H., Kasukawa, T., Saito, R.,
Kudoh, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Pletschmann, W., Gaasterland, T., Gissi, C., King, B., Kochwa, H.,
Knehl, P., Lewis, S., Matsuo, Y., Nikolaou, I., Peesele, G.,
Quackenbush, J., Schriml, L. M., Stabill, F., Suzuki, R., Tomita, M.,
Wagner, J. D., Wastli, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldirelli, P., Barsh, G., Blake, J., Hoffell, D., Bojunga, N.,
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,
Fleischer, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D. A., Kamuya, M., Lee, N. H., Lyons, P.,
Marchionni, L., Nishida, Y., Mazzarelli, J., Momtashim, P., Nodone, P.,
Rang, B., Rhinwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schoback, C., Sey, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyo-oka, K., Wang, K. H., Weltz, C., Whitteker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohzuki, S.,
and Hayashizaki, Y.

Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)

110855660

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

5 (bases 1 to 1040)

Adachi, Y., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C.,
Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T.,
Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiroaka, T., Holt, F.,
Hume, D., Imocani, K., Ishii, Y., Itoh, M., Iwata, K., Kasukawa, T.,
Kato, H., Kawachi, K., Kojima, Y., Konno, H., Kouda, M., Koya, S.,
Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K.,
Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,
Saito, H., Saito, R., Sakai, C., Sekai, K., Sano, H., Sasaki, D.,
Schmitt, L., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T.,
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
Tanaka, T., Tejima, Y., Toyai, T., Yamamura, T., Yamashita, I.,
Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and
Hayashizaki, Y.

TITLE
JOURNAL

Pharmacol. Ther. 2007; 105:149-160. © 2007 Tsunehiko Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-tes@gsc.riken.go.jp, UR: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9216, Fax: 81-45-503-9216).

COMMENT

For further details, please visit our web site (<http://genome.gsc.riken.go.jp/>)

FEATURES

Source

gene	1.1040	/gene="Rbpms"
CDS	266.	.928

BASE COUNT	223	a	352	c	247	g	217	t	1	others.
ORIGIN										
	<pre> /gene="Rbpms" /note="RNA binding protein gene with multiple splicing data source:MCD, source key:MGI:1334446, evidence:ISS putative" /codon_start=1 /protein_id="BAB28128.1" /db_xref="GI:12848898" /db_xref="MGI:1334446" /translation="MNGGCGAEEKENPSEANLOEEVRTLFGSLPIIDIKRPRELYLF RPFEGEGSLIKLTKSQVGFSPDSRSEAEAKNALNGIRPDEITPOTLRBEAKAN TKMAKNLTVGPNPSTPLTPVPOFIAREVELYTPALYSSPEWMAVPLYPETLAP ALPPPAFTYTPASHHACGCFSEAPKPNTPVCPPLQQIRFVSGNVEFYPTADQREL PC" </pre>									

Query Match	71.28;	Score 249.2;	DB 11;	Length 1040;
Best Local Similarity	79.18;	Pred. No. 3.7e-28;		
Matches 277;	Conservative 30;	Mismatches 38;	Indels 5;	Gaps 3

Dy 1 ATAGAGCTCACAGTACTGCACTTTACCCCGAGTAGCCTGAAGTGTCGGGCCCGTACCCTC 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 663 ATGAGCTCACAGTACTGCACTTTACCCCGAGTAGCCTGAAGTTGGGCCCGTACCCTC 722

61 TGTACCCAGCGAGTTAGCGCCCTGCTCTTCCCTCCTCGCGCTTTCACCTACCCGCTT 120

121 CACTGCATGCCGAGGCGCTGTATACATGCTAGGNNNNNTACNCGTACNNNNCTAACC 1000

783 CACTGCATGCCAGTGTTCCTCCAGAGGCAAGCCCAACACACACCTCTTTTGTCCAC 842

[illegible]

BY : 241 CAGACCACNCNDAGCCGACCGCTTTCATCAGCATTCGTTGGTGTGCATATATCAACTACTG 899
DB 843 TTTCTTCAGCA---GATTAGATTGTCTCTCTGGGAATGTGTTTGTCATATATCAACTACTG 899

Db 900 CAGACCAGCAGAGGAGCTCCCGTG- TTGAATTTATATAGCAACCCGCT- CCCCTCCCCCTT 952

301 TTTCCCAANGANAACCTNTTCTTTGANAATNTTTGAGAGATTTCAATAAAAA :350

Db 958 TTGCAAAAAAAAACTATTCTCTGATAATATTGAGAGCTTTCATTAATAAA 1007

RESULT 5
AK012586

LOCUS	1040 bp	mRNA	linear	HTC 19-JAN
AK012586				
DEFINITION	Mus.musculus 11 days embryo whole body cDNA, RIKEN full-length			

enriched library, clone:2700088L4:RNA binding protein gene with multiple splicing, full insert sequence.
AK012586

```

VERSION      AK012586.1  GI:12849425
KEYWORDS     HTC; CAP trapper.

```

mus musculus (strain: C57BL/6J) 11 days. embryo cDNA to mRNA,
clone-lib: RIKEN full-length enriched mouse cDNA library
clone: 3700088814

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi

REFERENCE
1. Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Carnivori; and Haverchivskiy

TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)

PubMed	10349636
REFERENCE	2

AUTHORS. Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.

Journal	Title
Genome Res 10 (10): 1617-1626 (2000)	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

MEDLINE · 20499374

REFERENCE 1 (bases 1 to 420)
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Stepien, M., Tan, F., Underwood, R., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
 TITLE The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNM; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:904588
 Seg primer: -40m3 fwd. ET from Amersham
 High quality sequence stop: 159.

FEATURES
 source
 1. 420
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:1382120"
 /clone_lib="Soares_mammary_gland_NMLMG"
 /sex="Female (lactating)"
 /tissue_type="mammary gland"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(df) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 116 a 71 c 132 g 101 t

ORIGIN

Query Match 69.8%; Score 244.4; DB 9; Length 420;
 Best Local Similarity 78.3%; Pred. No. 6.8e-28;
 Matches 274; Conservative 30; Mismatches 41; Indels 5; Gaps 3;

QY 1 ATAGGCTCAAGTACCTGCACTTACCCAGTACCTGAGTGTGGCCCCGTTACCTC 60
 DB 359 ATAGGCTCAAGTACCTGCACTTACCCAGTACCTGAGTGTGGCCCCGTTACCTC 300
 QY 61 TGTACCCAGGAGTATAGCCCTGCTCTCTCTCTGCGGCTTTACCTACCCGCTT 120
 DB 299 TGTACCCAGGAGTATAGCCCTGCTCTCTCTCTGCGGCTTTACCTACCCGCTT 240
 QY 121 CACTGATGCCAGAGGAGGCTGTATACATGCTAGGNNNNNACNCTGACNNNTGAGANAC 480
 DB 239 CACTGATGCCAGAGGAGGCTGTATACATGCTAGGNNNNNACNCTGACNNNTGAGANAC 480
 QY 181 TNCCTNACGACATGATGATGCGTCTGCTGGGATGCGCTGACACGATCAACTACTG 240
 DB 179 TTTCTTACGCA---GATTAGATTTGCTCTCTGGAATGTTGTCACATCAACTACTG 240
 QY 241 CACAGCANNAGAGGAGGCTGNNAGACAGATGATATTANNAACNACTACTATNCTT 300
 DB 122 CACAGCAGGAGGAGGAGGCTGNNAGACAGATGATATTANNAACNACTACTATNCTT 300
 QY 301 TTGCAAAAGNAACTTTCTTTGANAATNTTTAGAGACTTCAATAAAAA 350
 DB 64 TTGCAAAAGNAACTTTCTTTGANAATNTTTAGAGACTTCAATAAAAA 15

RESULT 7
 A0020489/c 423 bp mRNA linear EST 19-OCT-1998

DEFINITION A0020489 Mouse eight-cell stage embryo cDNA Mus musculus cDNA clone
 J0533602 3', mRNA sequence.
 ACCESSION A0020489
 VERSION A0020489.1 GI:3376073
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 423)
 AUTHORS Ko, M.S.H., Kitchen, J.R., Wang, X., Threalt, T.A., Sun, T., Depalma, G.E., Liang, Y., Kargul, G.J., Sharara, R. and Doi, H.
 TITLE Systematic analyses of genes expressed in eight-cell stage mouse embryos (The ERATO/Doi project at Wayne State University) (Ko, M.S.H. et al.)
 JOURNAL Unpublished (1998)
 COMMENT Contact: Hirotumi Doi
 Doi Biosymmetry Project, ERATO
 Japan Science and Technology Corporation (JST)
 WBS Marine East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan
 Email: hdo@bio.jst.go.jp.

FEATURES
 source
 1. 423
 Location/Qualifiers
 1. 423
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="J0533602"
 /clone_lib="Mouse eight-cell stage embryo cDNA"
 /dev_stage="Eight-cell stage embryo"
 /note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(df) primer [5']
 TGTACCAATCTGAACTGAGAGGAGGCGGAGATGTTTTTTTTTTTTTTTTTTT
 T 3'; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 119 a 73 c 132 g 98 t 1 others

ORIGIN

Query Match 69.0%; Score 241.6; DB 9; Length 423;
 Best Local Similarity 78.5%; Pred. No. 1.9e-27;
 Matches 270; Conservative 30; Mismatches 39; Indels 5; Gaps 3;

QY 7 TCAACGATGCTGACATTTACCCAGTACGCTGAGTGTGGCCCCGTTACCTCTGACC 66
 DB 352 TCAACGATGCTGACATTTACCCAGTACGCTGAGTGTGGCCCCGTTACCTCTGACC 293
 QY 67 CAGCGAGTTAGAGCGCTGCTCTCTCTCTGCGGCTTTACCTACCCGCTTACTGC 126
 DB 292 CAGCGAGTTAGAGCGCTGCTCTCTCTCTGCGGCTTTACCTACCCGCTTACTGC 223
 QY 127 ATGCCAGAGGAGGAGGCTGTATACATGCTAGGNNNNNACNCTGACNNNTGAGANACTGNC 166
 DB 232 ATGCCAGAGGAGGAGGCTGTATACATGCTAGGNNNNNACNCTGACNNNTGAGANACTGNC 166
 QY 187 AGCAGATGATGATGATGCGTCTGGAATGCGCTGACAGATCAACTACTGACAGACC 246
 DB 172 AGCA---GATTAGATTTGCTCTGGAATGTTGTCACATCAACTACTGACAGACC 116
 QY 247 ANCAAGAGGAGGCTGNNAGACAGATGATATTANNAACNAACTACTATNCTTTGCAA 306
 DB 115 AGCAGAGGAGGAGGCTGNNAGACAGATGATATTANNAACNAACTACTATNCTTTGCAA 58
 QY 307 ANCAAGATGATGATGATGCGTCTGGAATGCGCTGACAGATCAACTACTGACAGACC 350
 DB 57 AAAAAAAGTATTTCTTTGANAATNTTTAGAGACTTCAATAAAAA 14

RESULT 8

AU019051/c	AU019051	460 bp	mRNA	linear	EST 19-Oct-1998
LOCUS	AU019051	Mouse eight-cell stage embryo cDNA Mus musculus cDNA clone J0512C09.3'	mouse		
DEFINITION	J0512C09.3'	mRNA sequence.			
ACCESSION	AU019051				
VERSION	AU019051.1	G1:3374635			
KEYWORDS	EST.				
SOURCE	Mus musculus				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (cases 1 to 460)				
AUTHORS	Ko,M.S.H., Kitchen,J.R., Wang,X., Theault,T.A., Sun,T.T., DePalma,G.E., Liang,Y., Karquig,G.J., Sharrara,R. and Del,H.H.				
TITLE	Systematic analyses of genes expressed in eight-cell stage mouse embryos (The ERATO/Dot Project at Wayne State University) (Ko ,M.S.H. et al.)				
JOURNAL	unpublished (1998)				
COMMENT	Contact: Hirofumi Doi Dot Bioscience Project, ERATO Japan Science and Technology Corporation (JST) WBG Marine East 12F, 2-6 Nakase, Minama-Ku, Chiba 261-71, Japan Email: hdoibio@jst.go.jp. Location/Qualifiers 1..460 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone_id="J0512C09" /clone_lib="Mouse eight-cell stage embryo cDNA" /dev_stage="eight-cell stage embryo" /note="Organ: mammary gland; Vector: pET73D-Pac (pharmacia) with a modified polylinker; site_1: Not I; Site_2: Eco RI; lsc strand cDNA was primed with a Not I - oligo(dT) primer [5'. TCGTACCAGATCTGAAGTCGGAGCGCCGCGCATCGTTTCTTTTTTTTTTTT T 3-1]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Patima Bonaldo."				
BASE COUNT	120 a	80 c	143 g	116 t	1 others
ORIGIN					
Query Match	68.1%	Score 238.2;	Dk 9;	Length 460;	
Best Local Similarity	78.9%;	Pred. No. 7.4e-27			
Matches 277:	Conservative 30;	Mismatches 38;	Indels 6;	Gaps 4;	
QY	1 ATGACCTCACAGTACTTCACATTATGCCACATGCGCTG-AAGTGTGGCGCGGTACGCTT	59			
Dh	359 ATGAGCTCAAGTAAGTCTGACCTTTAGCCCAATAGCTTGAAAGTTTGGGCGCGGTACGCTT	300			
YY	60 CTGTACCGACGCAGATTAAGCGCTGCTCTTCCTGCTGCGTCTTCACGTACCCGCTT	119			
Dh	259 CTGTACCGACGCAGATTAAGCGCTGCTCTTCCTGCTGCGTCTTCACGTACCCGCTT	240			
QY	120 TCACGTACGTGCCCAGAGACGCTGTATACATGCTTAGGNNNNACNCGTACACNNNTGGANA	179			
Dh	239 TCACGTACGTGCCCAGAGACGCTGTATACATGCTTAGGNNNNACNCGTACACNNNTGGANA	180			
QY	180 CTGCTCAGACATCATGATGATGCTGCTGCTGCGAATGACCGCTGACGACATGACCTACT	239			
Dh	179 GTTCTTGAGCA---GATTAGATTTCTCTCTGGGAATGTCTTTGTACATATTCACCTACT	123			
QY	240 GCAGACGANCANAGAGGAGCTTNNAGACAGATGATTANNNAACGNACTACCTATATNCCT	299			
Dh	122 GCAGACGANCANAGAGGAGGAGCTCCCGTG-TTGAATTTATTAAGAACCCGCT-CGCTCCGCT	65			
QY	300 TTTTGGCAANGANAACTTTCTTTGGANANTTTTGAGACATTTGCAATAAAA	350			
Dh	64 TTTTGGCAAAAAAAACTATTTCTTTGATATTTATTTGAGACCTTTGCAATAAAA	14			

RESULT 9	BE945377/c	1607 bp	mRNA	linear	EST 03-oct-2000
LOCUS	BE945377				
DEFINITION	U1-M-MH3-avl1-c-01-0-UT.s1 NIH-BAP_M_S4 Mus musculus cDNA clone.				
ACCESSION	BE945377				
VERSION	BE945377.1				
KEYWORDS	BE945377.1 GI:10527136				
SOURCE	EST.				
ORGANISM	house mouse.				
	Mus musculus				
REFERENCE	Bukatyuta; Metazoa: Chordata: Cranialia: Vertebrata: Euteleostomi: Pannalia: Euthelia: Rodentia: Sciurognathi: Muridae: Murinae: Mus.				
AUTHORS	1 (bases 1 to 607)				
TITLE	Bonaldi, M.F., Lennon, G., and Soares, M.B.				
	Normalization and subtraction: two approaches to facilitate gene discovery				
JOURNAL	Genome Res.	6 (9),	791-806	(1996)	
MEDLINE	97044477				
COMMENT	Contact: Chiu, H				

JOURNAL: GENOME RES. 6 (9), 791-806 (1996)
MEDLINE: 9704447
COMMENT: Contract: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 5850
Email: m5stemal.nih.gov

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the brain stems tissue cDNA library preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMP6 cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMP6 cDNAs whose availability will be considered under appropriate and limited collaborative arrangements

Seq primer: M13 Forward
POLYA-Yes.

location/DnaI/inserts

l. :607

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="01-M-BH3-av1-c-01-01"

/clone_lib="NIH_BMAP_M_S4"

/dex_strage="27-32 days"

/lab_host="DH10B (Life Technologies)"

/note=Vector: pT73D-Pac (Pharmacia) with a modified polylinker Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_M_S4 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH_BMAP_M_S4, NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1, NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library (NIH_BMAP_M_S4) was constructed as follows: PCR-amplified cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the NIH_BMAP_M_S4 library. This procedure has been previously described (Donaldson Lennon and Soares, Genome Research

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: mscoares@blue.weeg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult heart library. cDNA library Preparation: M. Felina Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics This clone is also available through the I.M.A.G.E. Consortium at LLNL (info@imga.llnl.gov). IMAGE ID=1774034
Seq primer: M13 Forward
POLYA=NO.

ACCESSION	AI712872
VERSION	AI712872.1
KEYWORDS	GI:5016672
SOURCE	EST
	Norway rat.

ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Graniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 410)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dt track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonalde poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dt track served to verify it as a clone from the normalized ventricle at 13 dpc library cDNA library preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES
 source 1..410
 Location/Qualifiers
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="U1-R-AG1-aaf-e-04-0-01"
 /clone_lib="U1-R-AG1"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: p773D-Pac (Pharmacia) with a modified polylinker. Site.1: Not I; Site.2: Eco RI; The U1-R-AG1 library is a normalized library constructed from 13 dpc rat ventricle. The tag is a string of 6 nucleotides present between the Not I site and the oligo-dt track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Jim Lin, Department of Biology, University of Iowa.
 TAG_Lib=U1-R-AG1
 TAG_TISSUE=ventricle at 13 dpc
 TAG_SEQ=CACGCA"

BASE COUNT 111 a 72 c 120 g 107 t

ORIGIN

Query Match 64.3% Score 225.2; DB 9; Length 410;
 Best local Similarity 75.4%; Pred. No. 7; 5e-25;
 Matches 264; Conservative 29; Mismatches 48; Indels 9; Gaps 3;

QY 1 ATGAGCTCAGTACTGACTTTTACCCAGTAGAGCCCTGAGTGTGGCCCGCTACCTTC 60
 DB 371 ATGAGCTCAGTACTGACTTTTACCCAGTAGAGCCCTGAGTGTGGCCCGCTACCTTC 312

QY 61 TGTACCCAGGAGTTAGGACCTGCTCTTCCTCTCCCGGTTTCAGCTAACCGGCTT 120
 DB 311 TGTACCCAGGAGTTAGGACCTGCTCTTCCTCTCCCGGTTTCAGCTAACCGGCTT 252

QY 121 CACTGATCCCGGAGCGCGCTATACATGCGCTAGCGNNNNCACTGACNNNTGACANAC 180
 DB 251 CACTGATCCCGGAGCGCGCTATACATGCGCTAGCGNNNNCACTGACNNNTGACANAC 192

QY 181 TNCATGACCATGATATGCGCTGCGATATGCGCTGCGATATGCGCTGCGATATGCGCT 240
 DB 191 TNCATGACCATGATATGCGCTGCGATATGCGCTGCGATATGCGCTGCGATATGCGCT 135

QY 241 CAGACCAACCAAGGAGCTTNNGACAGCATGATATTANNAACCAAGCTTATGCTT 300
 |||||||: |||||||: ||| |||||||: ||| |||

DB 134 CAAACCAACAGAGGAGC-TCCATGTTGATTTATTAGCAAGCCTCCCTCCCTTT 76
 QY 301 TTGCAAAACAGACCTTTCTTGATATGTTTGGAGGATTTCAATAAAAA 350
 |||||||: |||||||: ||| |||||||: ||| |||

DB 75 T-----GCAAAAGCTATTTCTTGATGATATTTGAGAGATTCAATAAAA 31

RESULT 12
 B1288680/C
 LOCUS
 DEFINITION U1-R-DK0-cdd-f-11-0-01.s1 U1-R-DK0 Rattus norvegicus cDNA clone
 ACCESSION B1288680
 VERSION B1288680
 KEYWORDS B1288680.1 GI:14945502
 EST
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Graniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 410)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dt track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonalde poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dt track served to identify it as a clone from the normalized rat heart pool library cDNA library preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES
 source 1..410
 Location/Qualifiers
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="U1-R-DK0-cdd-f-11-0-01"
 /clone_lib="U1-R-DK0"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: p773D-Pac (Pharmacia) with a modified polylinker. Site.1: Not I; Site.2: Eco RI; The U1-R-DK0 library is a subtracted library derived from a mixture of five individually tagged normalized rat libraries: brain-nRBP (208), heart-nRBP (208), kidney-nRBP (208), liver-nRBP (208), and placenta-nRBP (208). Each original library was constructed from a mixture of equal amounts of RNA from seven different developmental time-points: embryonic day 17, embryonic day 19, embryonic day 21, adult day 1, adult day 12, adult day 75, and adult day 200. (Exception: the rat pool does not contain embryonic day 17 RNA and the placenta pool contains only the three embryonic stages). Each library was normalized individually according to the procedure described by Bonaldo, Lennon & Soares (Genome Research 6: 791-806, 1996). For construction of the DK0 subtracted library, plasmid DNA from each of the five individually tagged normalized libraries was mixed in the proportions specified above and electroporated into competent bacteria for production of single-stranded circular DNA representing the pool of libraries. Single-stranded


```
|||||
Db 252 CACTGCTATGCCAGTGTCTCTCCAGAGCAAGGCCACACACCGCTGTCTTGTCCAC 193
181 TCTNCAGCATGNATGATNGGTCGTCGGATGAGCTGCTGCAGCATGACATCTCTG 240
192 TCTTCCAGCAAT--TAGATTGTCTCTGGATGTGTTGTGCATTCACACCACTG 136
QY 241 CAGACGACNAGGAGCTTNGAGACAGATGTATPANNACGACTTATTCCTT 300
135 CAGACGACGAGAGGAGC-TCCCATGTTGAATTATTAGACAGCTCCCTGCCCTTT 77
Db 301 TTGCAANAGNAACTNTCTTTGANAATNTTTGAGAGATTTCATAAAA 350
76 T-----GCAAAACTATTCTGTATGATATTGAGAGATTTCATAAAAA 32

RESULT 14
A1103164/c 559 bp mRNA linear EST 31-JAN-1999
LOCUS A1103164
DEFINITION EST2.2453 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
ACCESSION A1103164
VERSION A1103164.1 GI:3707736
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 559)
AUTHORS Lee,N.H., Glodok,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@igf.org
Seq primer: M13-21.
FEATURES
source
1..559
/organism="Rattus sp."
/db_xref="ATCC (inproc):2024504"
/db_xref="taxon:10118"
/clone="RBMX96"
/clone_1lb="Normalized rat embryo, Bento Soares"
/dev_stage="embryo 8, 12, 18 dpc"
/note="Vector: pT7T3Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT 144 a 108 c 156 g 141 t
ORIGIN
Query Match 64.3%; Score 225.2; DB 9; Length 559;
Best Local Similarity 75.7%; Pred. No. 1.1e-24;
Matches 265; Conservative 28; Mismatches 48; Indels 9; Gaps 3;
Db 1 ATGAGCTCAGAGTACCTGCACTTACCCAGTAGCCCTGAAGTGTGGGCCCGTACCTC 60
354 ATGAGCTCAGAGTACCTGCACTTACCCAGTAGCCCTGAAGTGTGGGCCCGTACCTC 295
61 TGTAACGAGGAGATTAAGCGCTGCTCTCTCTCTGCGGCTTTCAGTACCGCGCTT 120
294 TGTAACGAGGAGATTAAGCGCTGCTCTCTCTCTGCGGCTTTCAGTACCGCGCTT 235
QY 121 CACTGCATGCCACAGAGCNCGTATACATGCTAGGNNNNACNCCCTAGCANN 180
135 CACTGCATGCCACAGAGCNCGTATACATGCTAGGNNNNACNCCCTAGCANN 175
Db 234 CACTGCATGCCACAGAGCNCGTATACATGCTAGGNNNNACNCCCTAGCANN 175
181 TCTNCAGCATGNATGATNGGTCGTCGGATGAGCTGCTGCAGCATGACATCTCTG 240
174 TCTTCCAGCAAT--TAGATTGTCTCTGGATGTGTTGTGCATTCACACCACTG 118
Db 174 TCTTCCAGCAAT--TAGATTGTCTCTGGATGTGTTGTGCATTCACACCACTG 118
```

```
QY 241 CAGACGACNAGGAGCTTNGAGACAGATGTATPANNACGACTTATTCCTT 300
117 CAGACGACGAGAGGAGCTCCCATGTTGAATTATTAGCAAGCT--TCCCTGCCCTT 60
QY 301 TTGCAANAGNAACTNTCTTTGANAATNTTTGAGAGATTTCATAAAA 350
59 TTGCA--AAACTATTCTGTATGATATTGAGAGATTTCATAAAAA 14

RESULT 15
A1231308 584 bp mRNA linear EST 31-JAN-1999
LOCUS A1231308
DEFINITION EST22795 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
ACCESSION A1231308
VERSION A1231308.1 GI:3815188
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 584)
AUTHORS Lee,N.H., Glodok,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@igf.org
Seq primer: M13-21.
FEATURES
source
1..584
/organism="Rattus sp."
/db_xref="ATCC (inproc):2038893"
/db_xref="taxon:10118"
/clone="RBM142"
/clone_1lb="Normalized rat embryo, Bento Soares"
/dev_stage="embryo 8, 12, 18 dpc"
/note="Vector: pT7T3Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT 147 a 114 c 174 g 145 t
ORIGIN
Query Match 64.3%; Score 225.2; DB 9; Length 584;
Best Local Similarity 75.4%; Pred. No. 1.2e-24;
Matches 264; Conservative 29; Mismatches 48; Indels 9; Gaps 3;
Db 1 ATGAGCTCAGAGTACCTGCACTTACCCAGTAGCCCTGAAGTGTGGGCCCGTACCTC 60
354 ATGAGCTCAGAGTACCTGCACTTACCCAGTAGCCCTGAAGTGTGGGCCCGTACCTC 295
61 TGTAACGAGGAGATTAAGCGCTGCTCTCTCTCTGCGGCTTTCAGTACCGCGCTT 120
294 TGTAACGAGGAGATTAAGCGCTGCTCTCTCTCTGCGGCTTTCAGTACCGCGCTT 235
QY 121 CACTGCATGCCACAGAGCNCGTATACATGCTAGGNNNNACNCCCTAGCANN 180
135 CACTGCATGCCACAGAGCNCGTATACATGCTAGGNNNNACNCCCTAGCANN 175
Db 234 CACTGCATGCCACAGAGCNCGTATACATGCTAGGNNNNACNCCCTAGCANN 175
181 TCTNCAGCATGNATGATNGGTCGTCGGATGAGCTGCTGCAGCATGACATCTCTG 240
174 TCTTCCAGCAAT--TAGATTGTCTCTGGATGTGTTGTGCATTCACACCACTG 118
QY 241 CAGACGACNAGGAGCTTNGAGACAGATGTATPANNACGACTTATTCCTT 300
117 CAGACGACGAGAGGAGC-TCCCATGTTGAATTATTAGCAAGCTCTCCCTGCCCTT 59
Db 117 CAGACGACGAGAGGAGC-TCCCATGTTGAATTATTAGCAAGCTCTCCCTGCCCTT 59
QY 301 TTGCAANAGNAACTNTCTTTGANAATNTTTGAGAGATTTCATAAAAA 350
```

Mon May 12 08:00:25 2003

us-09-880-711-328.rst

Page 12

Db 58 T-----GCAAAACCTATTCTTGATGATATTGAGAGATTTCATATAAAA 14

Search completed: May 9, 2003, 20:39:43
Job time : 1504 secs

GenCore version 5.1.4.p3.4578
Copyright (c) 1993 - 2003 Computer Ltd.

OM nucleic - nucleic search, using sw model

Run On: May 9, 2003, 19:38:48 ; Search time 67 Seconds

(without alignments)
1602.043 Million cell updates/sec

Title: US-09-880-711-328

Perfect score: 350
Sequence: 1 atgagcaccagctaccctca.....ctgaagatatttcaataaaa 150

Scoring table: IDENTITY_NUCDX

Gapop 10.0 , Gapext 1.0

Searched: 141362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_MA:*

- 1: /cgn2_6/pdata1/1/ina/5A_COMB.seq:*
- 2: /cgn2_6/pdata1/1/ina/5B_COMB.seq:*
- 3: /cgn2_6/pdata1/1/ina/6A_COMB.seq:*
- 4: /cgn2_6/pdata1/1/ina/6B_COMB.seq:*
- 5: /cgn2_6/pdata1/1/ina/PC105_COMB.seq:*
- 6: /cgn2_6/pdata1/1/ina/Backfiles...seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60.8	17.4	7218	1	US-08-232-463-14 Sequence 14, Appl
2	56.4	16.1	9060	4	US-08-378-313-20 Sequence 20, Appl
3	54.8	15.7	2035	4	US-08-560-780-10 Sequence 10, Appl
4	54.8	15.7	2035	4	US-09-073-898-10 Sequence 10, Appl
5	54.2	15.5	751	4	US-09-221-017B-18 Sequence 18, Appl
6	54.2	15.5	3619	4	US-08-377-503-1 Sequence 1, Appl
7	54.2	15.5	3619	4	US-08-178-019-1 Sequence 1, Appl
8	53.4	15.3	2200	1	US-08-592-125-102 Sequence 102, App
9	53.4	15.3	7452	3	US-08-592-500-1 Sequence 1, Appl
10	53.4	15.3	7452	3	US-08-195-005-1 Sequence 1, Appl
11	53.4	15.3	7452	5	US-08-936-155A-12 Sequence 12, Appl
12	52.4	15.0	1297	1	US-08-370-975B-12 Sequence 12, Appl
13	52.4	15.0	2593	4	US-08-936-155A-153 Sequence 153, App
14	52.4	15.0	5150	4	US-09-068-140A-14 Sequence 14, Appl
15	52.4	15.0	26764	1	US-08-370-975B-1 Sequence 1, Appl
16	52.2	14.9	1425	1	US-08-353-550-7 Sequence 7, Appl
17	52.2	14.9	1425	2	US-08-551-687-7 Sequence 7, Appl
18	52.2	14.9	1512	2	US-07-938-154-10 Sequence 10, Appl
19	52.2	14.9	1512	5	US-09-059-769-19 Sequence 19, Appl
20	52.2	14.9	1610	4	US-09-059-769-19 Sequence 19, Appl
21	52.2	14.9	1819	4	US-09-453-702B-188 Sequence 188, App
22	52.2	14.9	3031	4	US-08-664-962B-3 Sequence 3, Appl
23	52.2	14.9	3031	4	US-09-311-743-3 Sequence 3, Appl
24	51.6	14.7	3797	4	US-08-936-155A-107 Sequence 107, App
25	51.4	14.7	4853	4	US-08-881-450A-22 Sequence 22, Appl
26	51.2	14.6	3872	5	US-08-881-450A-22 Sequence 22, Appl
27	51.2	14.6	3116	1	US-08-149-103-2 Sequence 2, Appl

C 28	51	14.6	3116	1	US-08-451-883-2	Sequence 2, Appl
C 29	50.8	14.5	859	4	US-09-247-373B-47	Sequence 47, Appl
C 30	50.8	14.5	977	1	US-08-017-522A-1	Sequence 1, Appl
31	50.8	14.5	977	6	5215895-2	Patent No. 5215895
32	50.8	14.5	1100	1	US-07-949-516A-1	Sequence 1, Appl
33	50.8	14.5	1100	2	US-08-814-459-1	Sequence 1, Appl
34	50.8	14.5	1100	3	US-09-122-525-1	Sequence 1, Appl
35	50.8	14.5	2073	4	US-09-134-001C-1731	Sequence 1731, Ap
36	50.8	14.5	2167	3	US-08-368-704C-52	Sequence 52, Appl
37	50.8	14.5	2174	3	US-08-258-287B-54	Sequence 54, Appl
38	50.6	14.5	2119	4	US-09-240-519-7	Sequence 7, Appl
C 39	50.6	14.5	14104	4	US-08-961-527-34	Sequence 34, Appl
40	50.4	14.4	2175	2	US-08-389-564B-4	Sequence 4, Appl
41	50.4	14.4	2175	3	US-08-465-047B-4	Sequence 4, Appl
42	50.4	14.4	2437	1	US-07-795-859B-5	Sequence 5, Appl
43	50.4	14.4	2437	1	US-08-457-616-5	Sequence 5, Appl
44	50.4	14.4	2437	4	US-09-235-548-1	Sequence 1, Appl
C 45	50.4	14.4	8931	3	US-09-028-934-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: LAWRENCE, F.
APPLICANT: SCHIEPLINGER, F.
APPLICANT: FAULKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESS: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0295
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 1MMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZap1c-F15
US-08-232-463-14
Query Match: 17.4%; Score 60.8; DB 1; Length 7218;

Best Local Similarity 5.1%; Pred. No. 5.5;
Matches 17; Conservative 185; Mismatches 132; Indels 0; Gaps 0;

QY 6 CTCACAGTACCTGACCTTACCCGAGCCCTGAGTGTGGCCCCGCTGACCTGTAC 65
Db 1140 YY 1199
QY 66 CCAGCGAGTACGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 125
Db 1200 YY 1259
QY 126 CATGCCAGAGGCTGATACATGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 185
Db 1260 YY 1319
QY 186 CAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 245
Db 1320 YY 1379
QY 246 CANCAAGGAGCTTNNAGACAGATGATGATGATGATGATGATGATGATGAT 305
Db 1380 YY 1439
QY 306 AANGAAGTCT 339
Db 1440 AATCT 1473

RESULT 2

US-08-378-313-20/c
Sequence 20, Application US/08378313
Patent No. 620781
GENERAL INFORMATION:
APPLICANT: THEOLOGIS, ATHANASIOS
TITLE OF INVENTION: CONTROL OF FRUIT RIPENING THROUGH
TITLE OF INVENTION: GENETIC CONTROL OF ACC SYNTHASE SYNTHESIS
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOEISTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/378, 313
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/862,493
FILING DATE: 02-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 29190-20002.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 856-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 9060 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS

LOCATION: join(2704..2880, 2968..3099, 3183..3344, 3810
LOCATION: ..4376, 4463..4903)
US-08-378-313-20

Query Match 16.1%; Score 56.4; DB 4; Length 9060;
Best Local Similarity 40.7%; Pred. No. 51;
Matches 79; Conservative 29; Mismatches 86; Indels 0; Gaps 0;

QY 157 NNNACNCTGACNNNTGACNNCTGACNNCTGACNNCTGACNNCTGACNNCT 216
Db 7502 ATTTCACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7443
QY 217 NGCCTGACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 276
Db 7442 AGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7383
QY 277 TANNACNACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 336
Db 7382 TGCAACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 7323
QY 337 GATTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 350
Db 7322 AATTCAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7309

RESULT 3

US-08-960-780-10
Sequence 10, Application US/08960780
Patent No. 620435
GENERAL INFORMATION:
APPLICANT: Feltelson, Jerald S.
APPLICANT: Schepf, H. Ernest
APPLICANT: Narva, Kenneth E.
APPLICANT: Stockhoff, Brian A.
APPLICANT: Schmets, James
APPLICANT: Loewer, David
APPLICANT: Dullum, Charles Joseph
APPLICANT: Muller-Cohn, Judy
APPLICANT: Stamp, Lisa
TITLE OF INVENTION: Sequences which Encode These Toxins
NUMBER OF SEQUENCES: 134
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: US
ZIP: 32606-6669
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/960,780
FILING DATE: 30-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/029,848
FILING DATE: 30-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: MA-708
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: 158C2-pt1
US-08-960-780-10

Query Match
Best Local Similarity 40.3%; Pred. No. 15;
Matches 75; Conservative 29; Mismatches 82; Indels 0; Gaps 0;

15.7%; Score 54.8; DB 4; Length 2035;

QY 157 NNNACNCCGACGANNNGANACTNCTNCAGACATGATNATGCGTCNCTGGGAATG 216
Db 434 AAGTAAATCTGACCTGGCGAAGAAATCTCGACACGCTAGCTTATTAACGTAAATG 493
QY 217 NCCCTGNCAGATGACCTGACCTGACAGCANCNAGGAGCTTNNAGACGAAATGTAT 276
Db 454 TCGTATTAACTCTAGCGCTTACTGAAATTACACCTGCGGTATCAGCAATTAAATATCTGA 553
QY 277 TANNACGNACTACCTTATNCTTTGCAANGANAACTNTTCTTGANAATNTTTGAGA 336
Db 554 ATGAATAATTGACGATTTAAGCTTTTCTACAGAAACACTTTAAAGTAAAAAGGATA 613
QY 337 GATTTC 342
Db 614 GCTCTC 619

RESULT 4
US-09-073-898-10
Sequence 10, Application US/09073898
GENERAL INFORMATION:
APPLICANT: Fellefson, Jerald S.
APPLICANT: Schnepf, H. Ernest
APPLICANT: Narva, Kenneth E.
APPLICANT: Stockhoff, Brian A.
APPLICANT: Schmeltz, James
APPLICANT: Loewer, David
APPLICANT: Dullum, Charles Joseph
APPLICANT: Muller-Cohn, Judy
APPLICANT: Stamp, Lisa
APPLICANT: Mortill, George
APPLICANT: Finstad-Lee, Stacey
TITLE OF INVENTION: No. 624269e1 pesticidal toxins and Nucleotide
TITLE OF INVENTION: Sequences Which Encode These Toxins
NUMBER OF SEQUENCES: 144
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: US
ZIP: 32606-6669
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073.898
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/029,848
FILING DATE: 30-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/960,780
FILING DATE: 30-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Sanders, Jay M.
REGISTRATION NUMBER: 39,355
REFERENCE/DOCKET NUMBER: RA-708C1
TELECOMMUNICATION INFORMATION:

TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: 158C2-pt1
US-09-073-898-10

Query Match
Best Local Similarity 40.3%; Pred. No. 15;
Matches 75; Conservative 29; Mismatches 82; Indels 0; Gaps 0;

15.7%; Score 54.8; DB 4; Length 2035;

QY 157 NNNACNCCGACGANNNGANACTNCTNCAGACATGATNATGCGTCNCTGGGAATG 216
Db 434 AAGTAAATCTGACCTGGCGAAGAAATCTCGACACGCTAGCTTATTAACGTAAATG 493
QY 217 NCCCTGNCAGATGACCTGACCTGACAGCANCNAGGAGCTTNNAGACGAAATGTAT 276
Db 454 TCGTATTAACTCTAGCGCTTACTGAAATTACACCTGCGGTATCAGCAATTAAATATCTGA 553
QY 277 TANNACGNACTACCTTATNCTTTGCAANGANAACTNTTCTTGANAATNTTTGAGA 336
Db 554 ATGAATAATTGACGATTTAAGCTTTTCTACAGAAACACTTTAAAGTAAAAAGGATA 613
QY 337 GATTTC 342
Db 614 GCTCTC 619

RESULT 5
US-09-221-017B-18/1
Sequence 18, Application US/09221017B
Patent No. 6444795
GENERAL INFORMATION:
APPLICANT: ROSS, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221.017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1545
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Montoy, Gladys H.
REGISTRATION NUMBER: 32,430

RESULT 6
US-08-377-503-1/c
Sequence 1, Application US/08377503
Patent No. 632299
GENERAL INFORMATION:
APPLICANT: Kamboj, Rajender
APPLICANT: Elliot, Candace E.
APPLICANT: Nutt, Stephen L.
TITLE OF INVENTION: Kalnate-Binding, Human CNS Receptors of
TITLE OF INVENTION: the EAAs Family
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: Suite 500, 3000 K Street
City: N.W.
STATE: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/377,503
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/945,210
FILING DATE: 17-SEP-1992
ATTORNEY/AGENT INFORMATION:

US-08-178-019-1/C
Sequence 1, Application US/08178019
Patent No. 633161
GENERAL INFORMATION:
APPLICANT: Kamboj, Rajender
APPLICANT: Elliot, Candace E.
APPLICANT: Nutt, Stephen L.
TITLE OF INVENTION: Kalnate-Binding, Human CNS Receptors of
TITLE OF INVENTION: The EAA5 Family
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: Suite 500, 3000 K Street
CITY: N.W.
STATE: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy, disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/178,019
FILING DATE: 06-JAN-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Beutl, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16777/227
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3619 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 16..108
FEATURE:
NAME/KEY: mal-peptide
LOCATION: 109..2772
FEATURE:
NAME/KEY: CDS
LOCATION: 16..2772
US-08-178-019-1

Query Match 15.5%; Score 54.2; DB 4; Length 3619;
Best Local Similarity 40.1%; Pred. No. 41;
Matches 75; Conservative 29; Mismatches 83; Indels 0; Gaps 0;

QY 148 ATGCTAGNNNNCCAGCNCCTGACNNNTGAGNACTGTCAGCATGATGATNAGTCN 207
DB 1763 ATGACGACAGACGACAGACAGTACCCGACAGTAGCGCAGAGACATACATCGCATGTCT 1704
QY 208 CTGGAGATGCGCTGCGACAGATCACTACTGCGAGACCANCGAGGAGCTTTNCGACAC 267
DB 1703 GGGGACAGCGGATGAGGAGAGAGAGACGCTGGGCTTGCTCCATTGGCTTCGATAC 1644
QY 268 AGAATGATTANNACGNCNCTACTTATGCTTTTGCAANACNACTGTTGCTTGANAA 327
DB 1643 AGGATGCTACACGCAAGTCTGATGAAGGCTTGAGAGAGTGCATGCGCTTCTCGACACA 1584
QY 328 TTTTGA 334
DB 1583 TCGGTGA 1577

RESULT 8
US-08-592-126-102
Sequence 102, Application US/08592126
Patent No. 5821091
GENERAL INFORMATION:
APPLICANT: Gregory Doljancov
TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
NUMBER OF SEQUENCES: 151
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,126
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 2200 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: G89con.seq
US-08-592-126-102

Query Match 15.3%; Score 53.4; DB 1; Length 2200;
Best Local Similarity 39.1%; Pred. No. 30;
Matches 84; Conservative 30; Mismatches 101; Indels 0; Gaps 0;

QY 121 GACTGATGAGGAGAGGCGCTGATACATGCTAGGNNNNCCAGCNCCTGACNNNTGAGAC 180
DB 1817 CCCTGCTGAGCGGACAGAGCGCGGCTCTGCGCCGACAGCGCAACCTGAGAGCGCATCC 1876
QY 181 TWTGACGACATGATGATGATGCTGCTGCGGATGCGCTGCGACAGATGCAACTACTG 240
DB 1877 GCTTCTTACGACGACATGAGACGCAACGACGACCTTGATGCGCTTCTGATGCGCTTC 1935
QY 241 GAGACGACANCGAGGAGCTTNNAGACGACATGATATTANNACGNCNCTACTTATGCTT 300
DB 1937 CCAGCGCGGCGGCGGCGCCGACGAGACTGCTGCTGCACTTACGCGCGCTTCCCTTCTCG 1996
QY 301 TTGCAANACNACTTCTTGTGANAATTTTGAG 335
DB 1997 TTACATAGATACGACATGAGTCAAGCTCAGCGAGCGCTGAG 2031

RESULT 9
US-08-592-500-1
Sequence 1, Application US/08592500
Patent No. 6005085
GENERAL INFORMATION:
APPLICANT: Lanza, Francois
APPLICANT: Phillips, David R.
TITLE OF INVENTION: Gazezave, Jean-Pierre
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Kourie and Crew
STREET: 379 Lytton Avenue
CITY: Palo Alto
STATE: California
COUNTRY: US
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,500
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/089,455
FILING DATE: 09-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: How, Karen B.
REGISTRATION NUMBER: 29,684
REFERENCE/DOCKET NUMBER: 12418-28

```

1 NAME/KEY: misc_signal
2 LOCATION: 1365..1372
3 OTHER INFORMATION: /function="Ets-1 cis-acting
4 OTHER INFORMATION: sequences"
5 OTHER INFORMATION: /label=Ets-1
6 FEATURE:
7 NAME/KEY: repeat_region
8 LOCATION: 6133..6440
9 OTHER INFORMATION: /rpt_type="other"
10 OTHER INFORMATION: /label=Alu
11 FEATURE:
12 NAME/KEY: misc_signal
13 LOCATION: 5610..5615
14 OTHER INFORMATION: /standard_name="Polyadenylation
15 OTHER INFORMATION: signal sequence"
16 FEATURE:
17 NAME/KEY: misc_signal
18 LOCATION: 6966..6871
19 OTHER INFORMATION: /standard_name="Polyadenylation
20 OTHER INFORMATION: signal sequence"
21 FEATURE:
22 NAME/KEY: misc_signal
23 LOCATION: 7224..7229
24 OTHER INFORMATION: /standard_name="Polyadenylation
25 OTHER INFORMATION: signal sequence"
26 FEATURE:
27 NAME/KEY: misc_signal
28 LOCATION: 7358..7363
29 OTHER INFORMATION: /standard_name="Polyadenylation
30 OTHER INFORMATION: signal sequence"
31 FEATURE:
32 NAME/KEY: misc_feature
33 LOCATION: 1..7452
34 OTHER INFORMATION: /standard_name="Nucleotide
35 OTHER INFORMATION: sequence containing the human cpv gene"
36 US-08-592-500-1
37
38 Query Match 15.3%; Score 53.4; DB 3; Length 7452;
39 Best Local Similarity 39.2%; Fred. No. 1.5e+02;
40 Matches 78; Conservative 30; Mismatches 91; Indels 0; Gaps 0;
41
42 Oy 134 GAGGCGCTGTATACATGGCTAGGANNNCACNCCCTGACNNNTGGANACTGCTACAGACAT 193
43 |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
44 Db 6376 GTCGCACGACACTGTAGCTGGCGCACAGACGAGACATCTCAATCAAAAAAAGAAAAA 6435
45
46 Oy 134 GATATATGCGCTCCTGGAATGCGCTGNCACAGATCAACCTACTGACAGACCANMAG 253
47 |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
48 Db 6436 GAAATATCCAGTAATTAATAAATAATTGGGGAGAAATCTGATTTTACAAATACCTAG 6495
49
50 Oy 254 GAGCCTTNNAGACAGCATGTATTANNAACACACATCACTTATNCTTTTGGCAAGANNA 313
51 |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
52 Db 6496 TGTCTTCCGACGTAGCATGTCTTCATCTTCCCATTTATTAAGCATTTTAAATCTTTCA 6555
53
54 Oy 314 CTNTTCTTTGANAATNTTT 332
55 |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
56 Db 6556 GTGATGTTTAGAATTTT 6574
57
58 RESULT 10
59 US-08-195-006-1
60 Sequence 1, Application US/08195006
61 Patent No. 6083688
62 GENERAL INFORMATION:
63 APPLICANT: Lanza, Francois
64 APPLICANT: Phillips, David R.
65 APPLICANT: Cazenave, Jean-Pierre
66 TITLE OF INVENTION: Platelet glycoprotein V gene and uses
67 NUMBER OF SEQUENCES: 43
68 CORRESPONDENCE ADDRESS:
69 ADDRESSEE: Townsend and Townsend Kourile and Crew
70 STREET: 379 Lytton Avenue
71 CITY: Palo Alto
72 STATE: California

```



```

COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/195,005
FILING DATE: 10-FEB-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/089,455
FILING DATE: 09-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: DOV, Karen B.
REGISTRATION NUMBER: 25,664
REFERENCE/DOCKET NUMBER: 12418-28
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 326-2400
TELEFAX: (415) 326-2422
INFORMATION FOR SHO ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7452 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: intron
LOCATION: 1462..2419
FEATURE:
NAME/KEY: CDS
LOCATION: 2422..4101
FEATURE:
NAME/KEY: misc_signal
LOCATION: 68..76
OTHER INFORMATION: /function= "putative TPA responsive"
OTHER INFORMATION: /label= "element"
FEATURE:
NAME/KEY: misc_signal
LOCATION: 471..478
OTHER INFORMATION: /function= "Ets-1 cis-acting"
OTHER INFORMATION: /label= "Ets-1"
FEATURE:
NAME/KEY: misc_signal
LOCATION: 493..502
OTHER INFORMATION: /function= "Ets-1 cis-acting"
OTHER INFORMATION: /label= "Ets-1"
FEATURE:
NAME/KEY: repeat_region
LOCATION: 593..881
OTHER INFORMATION: /rpt_type= "other"
OTHER INFORMATION: /label= "Alu"
FEATURE:
NAME/KEY: misc_signal
LOCATION: 897..904
OTHER INFORMATION: /function= "Ets-1 cis-acting"
OTHER INFORMATION: /label= "Ets-1"
FEATURE:
NAME/KEY: misc_binding
LOCATION: 1142..1149
OTHER INFORMATION: /function= "Sp1 binding site"
OTHER INFORMATION: /standard_name= "sp1"
FEATURE:
NAME/KEY: misc_signal
LOCATION: 1178..1184

```

Query Match	Best Local Similarity	Matches	Conservative	Score	DB	Length
134	39.2%	78	30	53.4	3	7452
6376	39.2%	78	30	53.4	3	7452
194	39.2%	78	30	53.4	3	7452
6436	39.2%	78	30	53.4	3	7452
254	39.2%	78	30	53.4	3	7452
6496	39.2%	78	30	53.4	3	7452
314	39.2%	78	30	53.4	3	7452

Db 6556 GTGATGTTTAGAATTTT 6574

RESULT 11

PCT-US94-07644A-1

Sequence 1, Application PC/TUS9407644A

GENERAL INFORMATION:

APPLICANT: COR Therapeutics, Inc.

TITLE OF INVENTION: Platelet Glycoprotein V Gene and Uses

NUMBER OF SEQUENCES: 43

CORRESPONDENCE ADDRESS:

STREET: 379 Lytton Avenue

CITY: Palo Alto

STATE: California

COUNTRY: US

ZIP: 94301

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/07644A

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Dow, Karen B.

REGISTRATION NUMBER: 29,684

REFERENCE/DOCKET NUMBER: 012418-003000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 326-2400

TELEFAX: (415) 326-2422

INFORMATION FOR SEQ. ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 7452 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: intron

LOCATION: 1462..2419

FEATURE:

NAME/KEY: CDS

LOCATION: 2422..4101

FEATURE:

NAME/KEY: misc_signal

LOCATION: 68..76

FEATURE:

OTHER INFORMATION: /function= "Putative TPA responsive"

OTHER INFORMATION: /label= TRE

FEATURE:

NAME/KEY: misc_signal

LOCATION: 471..478

FEATURE:

OTHER INFORMATION: /function= "Ets-1 cis-acting"

OTHER INFORMATION: /label= Ets-1

FEATURE:

NAME/KEY: repeat_region

LOCATION: 593..861

FEATURE:

OTHER INFORMATION: /rpt_type= "other"

OTHER INFORMATION: /label= Alu

FEATURE:

NAME/KEY: misc_signal

LOCATION: 897..904

OTHER INFORMATION: /function= "Ets-1 cis-acting"

OTHER INFORMATION: /label= Ets-1

FEATURE:

NAME/KEY: misc_binding

LOCATION: 1142..1149

FEATURE:

OTHER INFORMATION: /function= "Sp1 binding site"

OTHER INFORMATION: /standard_name= "Sp1"

FEATURE:

NAME/KEY: misc_signal

LOCATION: 1178..1184

FEATURE:

OTHER INFORMATION: /function= "Ets-1 cis-acting"

OTHER INFORMATION: /label= Ets-1

FEATURE:

NAME/KEY: TATA_signal

LOCATION: 1199..1203

FEATURE:

NAME/KEY: TATA_signal

LOCATION: 1263..1269

FEATURE:

NAME/KEY: misc_binding

LOCATION: 1285..1289

FEATURE:

OTHER INFORMATION: /function= "GATA-1 binding site"

OTHER INFORMATION: /label= GATA-1 binding site"

FEATURE:

NAME/KEY: misc_signal

LOCATION: 1321..1326

FEATURE:

OTHER INFORMATION: /function= "GATA-1 binding site"

OTHER INFORMATION: /label= GATA-1 binding site"

FEATURE:

NAME/KEY: misc_signal

LOCATION: 1365..1372

FEATURE:

OTHER INFORMATION: /function= "Ets-1 cis-acting"

OTHER INFORMATION: /label= Ets-1

FEATURE:

NAME/KEY: repeat_region

LOCATION: 6133..6440

FEATURE:

OTHER INFORMATION: /rpt_type= "other"

OTHER INFORMATION: /label= Alu

FEATURE:

NAME/KEY: misc_signal

LOCATION: 5610..5615

FEATURE:

OTHER INFORMATION: /standard_name= "Polyadenylation"

OTHER INFORMATION: /label= Polyadenylation

FEATURE:

NAME/KEY: misc_signal

LOCATION: 7224..7229

FEATURE:

OTHER INFORMATION: /standard_name= "Polyadenylation"

OTHER INFORMATION: /label= Polyadenylation

FEATURE:

NAME/KEY: misc_signal

LOCATION: 7358..7363

FEATURE:

OTHER INFORMATION: /standard_name= "Polyadenylation"

OTHER INFORMATION: /label= Polyadenylation

FEATURE:

NAME/KEY: misc_signal

LOCATION: 1..7452

FEATURE:

OTHER INFORMATION: /standard_name= "Nucleotide"

PCT-US94-07644A-1

Query Match

Best Local Similarity 39.2%; Score 53.4; DB 5; Length 7452;

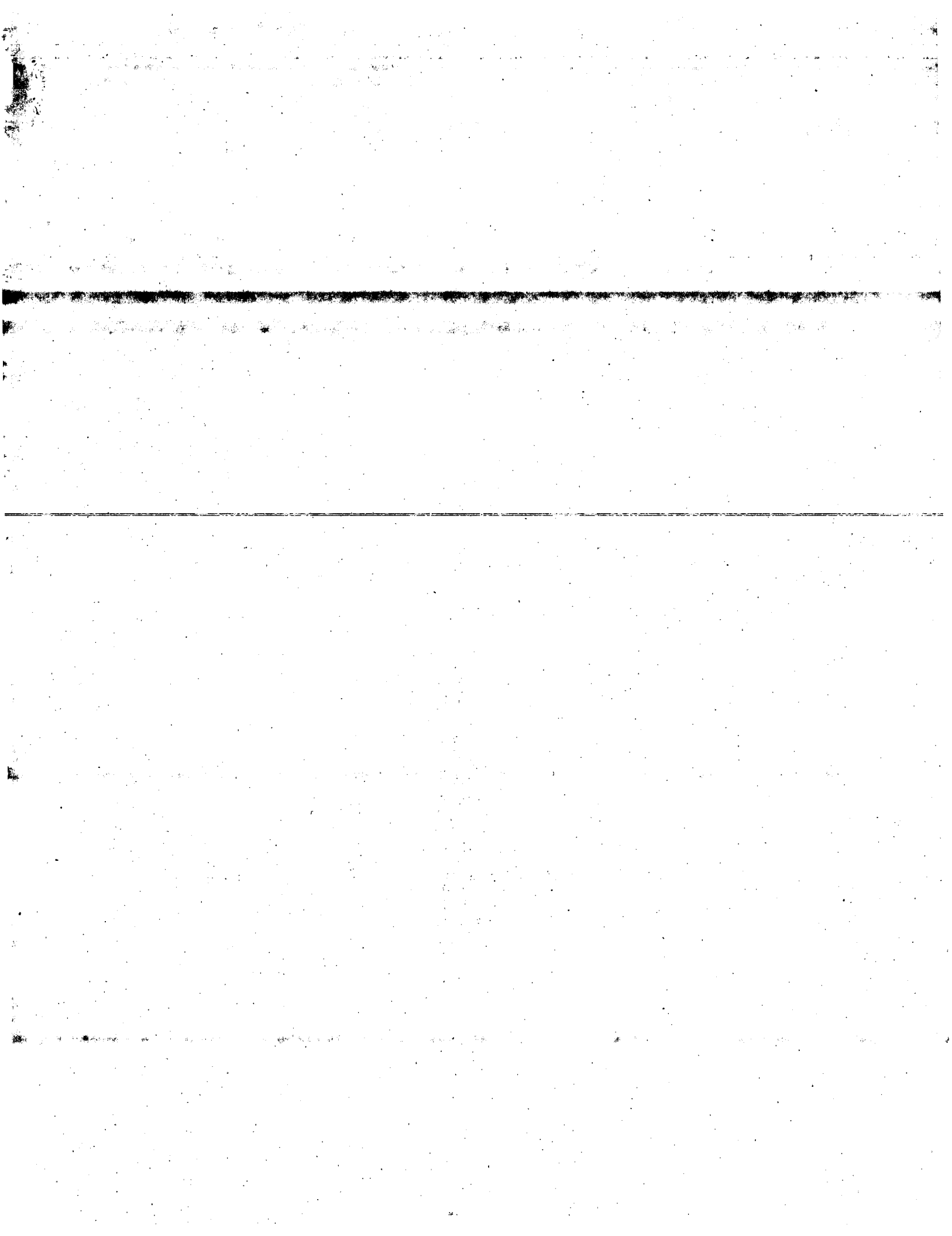
Matches 78; Conservative 30; Mismatches 91; Indels 0; Gaps 0;

QY 134 GAGGCMCTGATACATGCTAGCCGNNNNNACNCTGACNNNNNGAAGACTGCTACACACAT 193

DB 6376 GTGCCACTGCATCTAGCCTGGGCGACGAGCAGACTCCATCTCAAAAAAAGAAAAA 6435

QY 317 TTCTTGGANAATNTTGA 334
| ||| : | : ||| |
Db 26660 TATTTTCATGAGTTTAA 26677

Search completed: May 9, 2003, 20:41:29
Job time : 99 secs



Db 3011 GCCGAGATGTTCCAAATCTCTGACCTCGATGCCCATGGTGGCCCTCCCAAAATGCT 2952
 Oy 140 CTGATATCAATCCCTAGGNNNNKACNCCCTACANNNTGGAAACTNCTNCAGCACATGATGA 199
 Db 2951 GGGATTTCAGAGCGTGAAGCCACCATGCCCGCCCTTAGCAAAATATTTTACACACTATTTTCA 28922
 Oy 200 TNGCGTCNCTGGGAATGNGCCCTGNCACAGATCAACCTACTGCAGACCANCNAAAGGAGCT 259
 Db 2891 TCTTTTATTTTACAGCAAACTGAGAGGTGAGCGAGTCAAGCGAGATTTATTTATTTCA 28322
 Oy 260 TNNAGCACAAATGTGTTANNNAACACAGTACTTATNCTTTTGAAGNANNAACTNCTC 319
 Db 2831 TTTAATCAAAATTTGTAATAATATACATAACCATTTTCTGAATGTCTGTGAAGTAACTTTGGCT 27722
 Oy 320 TTTGANAATNTTT 332
 Db 2771 TTGCATGATCTAT 2759

RESULT 4
 US-09-764-877-33322/c
 Sequence 3332. Amplification HS/09764877

```

RESULT 4
US-09-764-877-3322/c
; Sequence 3322; Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antipodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; PRIORITY FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3322
; LENGTH: 8677
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-3322

Query Match
Best local similarity 40.2%; Pred. No. 1.9e+02;
Matches 84; Conservative 30; Mismatches 95; Indels 0; Gaps 0;

16.3%; Score 57; DB 10; Length 8677;

QY 121 CACTGCATGCCCGAGAGGCGCTGTATACATGCTTAGNNNNACNCCGTGACNNNTGGANAC 180
||||| | ||| | : |||| | |||| | : ||||| : |
4977 CACTAGCGCCCTACAGAGAGCTGTATACACTGAGCCCGCAGTGCAGTGAGAGCACTTGTGGGGTTC 4918

```

RESULT 5
 US-09-764-877-3321/C
 ; Sequence 3321, Application US/09764877
 ; Patent No. US2020147140A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

```
FILE REFERENCE: F000
CURRENT APPLICATION NUMBER: US/09/764,877
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 4031
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3321
```


LENGTH: 8680
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-877-3321

Query Match 16.3% Score 57 DB 10 Length 8680

Best Local Similarity 40.2% Pred. No. 1.9e+02

Matches 84: Conservative 30; Mismatches 95; Indels 0; Gaps 0;

QY 121 CACTGACATGGCCAGAGCGCTGTATACCTGGCTTACAGNNNNNCAACCTGACNNNTGCAAC 180
DB 4979 CACTGACGCTTACAGAGCGCTGTATACCTGACGCCCACTGACGAGGAGGAGCTTGGGGCTC 4920
QY 181 TNCNCAAGCATGATGATGCTGCTGCAATGNCCTGCAACATGACATGACCTACTG 240
DB 4919 ACCTGATCGGGCCCTGCTGCTGCGCAGGGCTGTGAGCTCCGACAGCAAGAGCCATTTCTGCTC 4860
QY 241 CAGACCCAGCAGGAGCTTNNAGACAGATGTATTANMAACCACTTATTCCTT 300
DB 4859 AGGACCCAGCAGGAGCTTNNAGACAGATGTATTGCTGACGCTGCTTACTGCTGAGC 4800
QY 301 TTGCAANGACAACTTCTTGANMATN 329
DB 4799 TGACTGCTGCCCGCTTTGCTTGATGTG 4771

RESULT 6

US-09-938-842A-755

Sequence 765, Application US/09938842A

Patent No. US20020160378A1

GENERAL INFORMATION:

APPLICANT: Harper, Jeff

APPLICANT: Kreps, Joel

APPLICANT: Wang, Xun

APPLICANT: Zhu, Tong

TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

TITLE OF INVENTION: SAME, AND METHODS OF USE

FILE REFERENCE: S01P300-3

CURRENT APPLICATION NUMBER: US/09/938, 842A

CURRENT FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: US 60/227, 866

PRIOR FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: US 60/264, 647

PRIOR FILING DATE: 2001-01-16

PRIOR APPLICATION NUMBER: US 60/300, 111

PRIOR FILING DATE: 2001-06-22

NUMBER OF SEQ ID NOS: 5379

SEQ ID NO 765

LENGTH: 621

TYPE: DNA

ORGANISM: Arabidopsis thaliana

US-09-938-842A-765

Query Match 16.1% Score 56.4 DB 9 Length 621

Best Local Similarity 39.7% Pred. No. 6.9

Matches 96: Conservative 30; Mismatches 116; Indels 0; Gaps 0;

QY 88 TTGCTGCTGCTGCGCGCTTACCTACCGCGCTTCACTGCAATGCGCAGAGCGCTGTATAC 147
DB 134 TCCCTGCTGCACTTCTTGACCTCTGCGCGCTTGTGCTCATGCTGAGCTTTAAGTACA 193
QY 148 ATGCTAGGNNNCAACCTGACNNNTGANAACCTGCTGAGTATGATGCTGCTG 207
DB 194 ACATGAGGCTGAGAACCGGTAAAGGATTCACCTTGAAGAGGTCAAGGCTGTGCTATGC 253
QY 208 CTGGGATGCGCTGNCAGATCACTACTGACAGACCAANMAAGGAGCTTNNAGAC 267
DB 254 CAAAGCACTTGGCGCTGCAATGCTATTGCTGTGACCATGCTGCAAGCAAGCATCTT 313
QY 268 AGAATGATTANNACCACTACTTATNCCTTTGCAANJANNAAGCTTCTTGANNA 327
DB 314 TCGAGGCTTTCACAGCAATGCTGACAGGCTCAAAACCTACAGACCAAGTACTTATTT 373

QY 328 TN 329
DB 374 TC 375

RESULT 7

US-09-770-445-733

Sequence 733, Application US/09770445

Patent No. US20020023281A1

GENERAL INFORMATION:

APPLICANT: Gottlieb, Jörn

APPLICANT: An, Yong-Oilang

APPLICANT: Hamilton, Carol M.

APPLICANT: Price, Jennifer L.

APPLICANT: Raines, Tracy M.

APPLICANT: Yu, Yang

APPLICANT: Kaneaka, Joshua G.

APPLICANT: Page, Amy

APPLICANT: Mathew, Abraham V.

APPLICANT: Ledford, Brooke L.

APPLICANT: Moessner, Jeffrey P.

APPLICANT: Haas, William David

APPLICANT: Garcia, Carlos A.

APPLICANT: Krieger, MaJa

APPLICANT: Slader, Ted

APPLICANT: Davis, Keith R.

APPLICANT: Allen, Keith

APPLICANT: Hoffman, Neil

APPLICANT: Burhan, Patrick

TITLE OF INVENTION: Expressed sequences of Arabidopsis

TITLE OF INVENTION: thaliana

FILE REFERENCE: 202305 (PAPA-012PRV)

CURRENT APPLICATION NUMBER: US/09/770, 445

CURRENT FILING DATE: 2001-01-25

PRIOR APPLICATION NUMBER: US 60/178, 472

PRIOR FILING DATE: 2000-01-27

NUMBER OF SEQ ID NOS: 999

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 733

LENGTH: 820

TYPE: DNA

ORGANISM: Arabidopsis thaliana

US-09-770-445-733

Query Match 16.1% Score 56.4 DB 10 Length 820

Best Local Similarity 39.7% Pred. No. 10

Matches 96: Conservative 30; Mismatches 116; Indels 0; Gaps 0;

QY 88 TTGCTGCTGCTGCGCGCTTACCTACCGCGCTTCACTGCAATGCGCAGAGCGCTGTATAC 147
DB 169 TCCCTGCTGCACTTCTTGACCTCTGCGCGCTTGTGCTCATGCTGAGCTTTAAGTACA 248
QY 148 ATGCTAGGNNNCAACCTGACNNNTGANAACCTGCTGAGTATGATGCTGCTG 207
DB 249 ACATGAGGCTGAGAACCGGTAAAGGATTCACCTTGAAGAGGTCAAGGCTGTGCTATGC 308
QY 208 CTGGGATGCGCTGNCAGATCACTACTGACAGACCAANMAAGGAGCTTNNAGAC 267
DB 309 CAAAGCACTTGGCGCTGCAATGCTATTGCTGTGACCATGCTGCAAGCAAGCATCTT 368
QY 268 AGAATGATTANNACCACTACTTATNCCTTTGCAANJANNAAGCTTCTTGANNA 327
DB 369 TCGAGGCTTTCACAGCAATGCTGACAGGCTCAAAACCTACAGACCAAGTACTTATTT 428
QY 328 TN 329
DB 429 TC 430

RESULT 8

US-09-764-877-3171/C

Sequence 3171, Application US/09764877

Patent No. US20020147140A1

```

1      PRIOR APPLICATION NUMBER: 09/774,636
2      PRIOR FILING DATE: 2001-02-01
3      PRIOR APPLICATION NUMBER: 60/238,291
4      PRIOR FILING DATE: 2000-10-06
5      PRIOR APPLICATION NUMBER: 09/244,111
6      PRIOR FILING DATE: 1999-02-04
7      PRIOR APPLICATION NUMBER: PCT/US98/
8      PRIOR FILING DATE: 1998-08-04
9      PRIOR APPLICATION NUMBER: 60/056,37
10     PRIOR FILING DATE: 1997-08-19
11     PRIOR APPLICATION NUMBER: 60/056,73
12     PRIOR FILING DATE: 1997-08-19
13     PRIOR APPLICATION NUMBER: 60/056,36
14     PRIOR FILING DATE: 1997-08-19
15     PRIOR APPLICATION NUMBER: 60/056,36
16     PRIOR FILING DATE: 1997-08-19
17     PRIOR APPLICATION NUMBER: 60/056,73
18     PRIOR FILING DATE: 1997-08-19
19     PRIOR APPLICATION NUMBER: 60/056,55
20     PRIOR FILING DATE: 1997-08-19
21     PRIOR APPLICATION NUMBER: 60/056,56
22     PRIOR FILING DATE: 1997-08-19
23     PRIOR APPLICATION NUMBER: 60/055,97
24     PRIOR FILING DATE: 1997-08-18
25     PRIOR APPLICATION NUMBER: 60/055,98
26     PRIOR FILING DATE: 1997-08-18
27     PRIOR APPLICATION NUMBER: 60/055,31
28     PRIOR FILING DATE: 1997-08-05
29     PRIOR APPLICATION NUMBER: 60/054,80
30     PRIOR FILING DATE: 1997-08-05
31     PRIOR APPLICATION NUMBER: 60/054,80
32     PRIOR FILING DATE: 1997-08-05
33     PRIOR APPLICATION NUMBER: 60/054,80
34     PRIOR FILING DATE: 1997-08-05
35     PRIOR APPLICATION NUMBER: 60/054,80
36     PRIOR FILING DATE: 1997-08-05
37     PRIOR APPLICATION NUMBER: 60/054,80
38     PRIOR FILING DATE: 1997-08-05
39     PRIOR APPLICATION NUMBER: 60/054,80
40     PRIOR FILING DATE: 1997-08-05
41     PRIOR APPLICATION NUMBER: 60/054,80
42     PRIOR FILING DATE: 1997-08-05
43     PRIOR APPLICATION NUMBER: 60/054,80
44     PRIOR FILING DATE: 1997-08-05
45     PRIOR APPLICATION NUMBER: 60/054,80
46     PRIOR FILING DATE: 1997-08-05
47     PRIOR APPLICATION NUMBER: 60/054,80
48     PRIOR FILING DATE: 1997-08-05
49     PRIOR APPLICATION NUMBER: 60/054,80
50     PRIOR FILING DATE: 1997-08-05
51     PRIOR APPLICATION NUMBER: 60/054,80
52     PRIOR FILING DATE: 1997-08-05
53     PRIOR APPLICATION NUMBER: 60/054,80
54     PRIOR FILING DATE: 1997-08-05
55     PRIOR APPLICATION NUMBER: 60/054,80
56     PRIOR FILING DATE: 1997-08-05
57     PRIOR APPLICATION NUMBER: 60/054,80
58     PRIOR FILING DATE: 1997-08-05
59     PRIOR APPLICATION NUMBER: 60/054,80
60     PRIOR FILING DATE: 1997-08-05
61     PRIOR APPLICATION NUMBER: 60/054,80
62     PRIOR FILING DATE: 1997-08-05
63     PRIOR APPLICATION NUMBER: 60/054,80
64     PRIOR FILING DATE: 1997-08-05
65     PRIOR APPLICATION NUMBER: 60/054,80
66     PRIOR FILING DATE: 1997-08-05
67     PRIOR APPLICATION NUMBER: 60/054,80
68     PRIOR FILING DATE: 1997-08-05
69     PRIOR APPLICATION NUMBER: 60/054,80
70     PRIOR FILING DATE: 1997-08-05
71     PRIOR APPLICATION NUMBER: 60/054,80
72     PRIOR FILING DATE: 1997-08-05
73     PRIOR APPLICATION NUMBER: 60/054,80
74     PRIOR FILING DATE: 1997-08-05
75     PRIOR APPLICATION NUMBER: 60/054,80
76     PRIOR FILING DATE: 1997-08-05
77     PRIOR APPLICATION NUMBER: 60/054,80
78     PRIOR FILING DATE: 1997-08-05
79     PRIOR APPLICATION NUMBER: 60/054,80
80     PRIOR FILING DATE: 1997-08-05
81     PRIOR APPLICATION NUMBER: 60/054,80
82     PRIOR FILING DATE: 1997-08-05
83     PRIOR APPLICATION NUMBER: 60/054,80
84     PRIOR FILING DATE: 1997-08-05
85     PRIOR APPLICATION NUMBER: 60/054,80
86     PRIOR FILING DATE: 1997-08-05
87     PRIOR APPLICATION NUMBER: 60/054,80
88     PRIOR FILING DATE: 1997-08-05
89     PRIOR APPLICATION NUMBER: 60/054,80
90     PRIOR FILING DATE: 1997-08-05
91     PRIOR APPLICATION NUMBER: 60/054,80
92     PRIOR FILING DATE: 1997-08-05
93     PRIOR APPLICATION NUMBER: 60/054,80
94     PRIOR FILING DATE: 1997-08-05
95     PRIOR APPLICATION NUMBER: 60/054,80
96     PRIOR FILING DATE: 1997-08-05
97     PRIOR APPLICATION NUMBER: 60/054,80
98     PRIOR FILING DATE: 1997-08-05
99     PRIOR APPLICATION NUMBER: 60/054,80
100    PRIOR FILING DATE: 1997-08-05
101    NUMBER OF SEQ ID NOS: 373
102    SOFTWARE: PatentIn Ver. 2.0
103    SEQ ID NO: 92

```


Query Match	15.7%;	Score 54.8;	DB 9;	Length 513509;
Best Local Similarity	39.4%;	Pred. No. 2.3e+03;		
Matches	86;	Conservative	30;	Mismatches 102;
			Indels	0;
			Gaps	0;

Db 413679 AGACTAAAGAAGAAATTTATTAAACACACACTACACCTACAAAGAGTGTCTTGATGCTTA 4137388
OY 133 AGAGCGCMCTGATTCATCAATGCCCTAGGNNNNNACNCTGCACNNNTGGANAATCTTCNACAGACA 192
Db 413679 AGACTAAAGAAGAAATTTATTAAACACACACTACACCTACAAAGAGTGTCTTGATGCTTA 4137388
OY 193 TGNMTGATNGGCTGCTGCGGAAATGCGCTGTCNCAAGATCAACCTACAGCAGACACANCMNA 252
Db 413739 ATGAGATATACATCATTTTAGAAGAGTAAACTAAAGAAGAGCATTTAATGCACTAACACACT 4137988

[illegible]

QY 313 ACTNTTCCTTGANAATNTTGGAGAGATTTCATAAAAA 350
| : | | : | | | | | | | |
Db 413859 AGATGTTTGCTGCTATATTATATATATATATATA 413896

RESULT 15
US-09-960-352-9955
: Sequence 9955, Application US/09960352
: Patent No. US20020137139A1

```

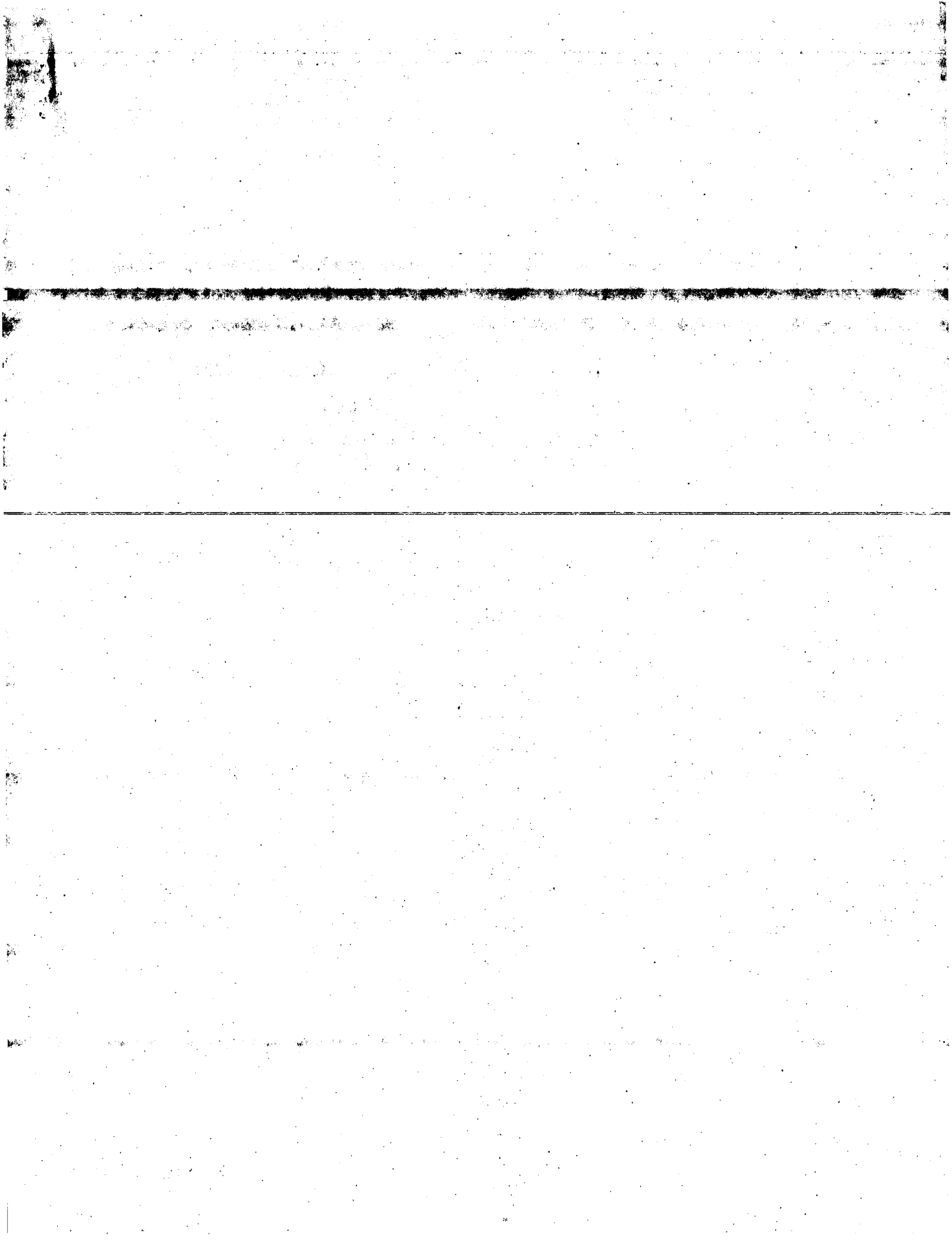
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Ningbing
APPLICANT: Wyatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 9955
LENGTH: 363
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 43-LIB2809-015-Q1-E1-C4
US-09-960-352-9955

```

Query Match:	15.68;	Score 54.6;	DB 10;	Length 363;
Best Local Similarity:	42.18;	Pred. No. 7.3;		
Matches 98;	Conservative 30;	Mismatches 104;	Indels 1;	Gaps
QY	105	TTCACTACCCCGCTTCACTGCAATGCCAGGCGCTGTATACCTGCTAGNNNNACN	164	
Db	63	TTGAGAGCGGAGTTACTATGTTAGACTGGGATCTCTGCCACAACTCGCCTCAAGGG	122	
QY	165	CGTGCANNNTGGANACTCTNCAGACATGNATGATNGCGTCTCNCCTGGGAATGMCCTGNC	224	
Db	123	CGTACCGCAGCGCCCTTTGTAGGGTTGAAGAGCTAAGCG-AAAAGCGCCAGGAGACCATTT	181	
QY	225	ACAGATCAACCTACTGCGAGCCANCAAGGAGCTTTNGAGACAGATGTATTTANNAACC	284	

Db 182 TCTCAGCTTCACFTCCGCCCTTCAACCTGAGTGAACCTTGGCCAGAGAAATGTGCATATATGCC 241
QY 285 NAGTACCTTATNGCTTTTGCAGANGANAGACTNTTCTTTGANAATNTTTGAGAG 337
Db 242 AACGAGAAAATTGAGATGCTGAGATAGCTATATCTGTCTCGGCTGGAGTG 294

Search completed: May 9, 2003, 20:47:52
Job time : 402 secs



Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES						
Result NO.	Score	Query Match	Length	DB	ID	Description
1	132	24.8	370	22	AA664541	Genetic hydrophobin
2	131	24.6	158	22	AAU30663	Novel human secret
3	131	24.6	730	18	AAW13888	Genetic formula #
4	131	25.9	730	18	AAW13888	Genetic formula #
5	130	25.7	119	22	AAU07680	Novel human diagen
6	129	24.2	731	18	AAW13871	Genetic formula #
7	129	25.5	731	18	AAW13871	Genetic formula #
8	126	24.2	731	21	AAW13807	Genetic formula of
9	125	25.5	731	21	AAW13807	Genetic formula of
10	128	25.3	411	21	AAW79913	Plasmodium DBL gen
11	127	23.8	440	23	AAU021480	Consensus protein
12	126	24.9	740	19	AAW37150	Murine Mena, Ena-V
13	125.5	23.5	155	21	AAV49970	Hydrophobic generi
14	125	24.7	370	22	AA664541	Genetic hydrophobin
15	124	23.3	176	21	AAV79585	EGF-like domain co
16	124	23.3	176	23	AAW17038	Human G protein-co
17	124	23.3	357	20	AAV37567	C. trachomatis prot
18	123	24.3	126	21	AAW48654	zevadi growth fact
19	123	24.3	134	22	AAU30506	Novel human secre
20	123	24.3	175	21	AAV79385	EGF-like domain co
21	123	24.3	176	23	AAW17038	Human G protein-co
22	123	23.1	261	21	AAW84325	Consensus sequence
23	123	24.3	328	14	AAW37615	Sequence of select
24	123	24.3	442	19	AAW86283	L. infantum cycle
25	123	24.3	442	20	AAV43539	A. leishmania infar
26	122	24.1	116	21	AAV64859	Cyteline knoi cons
27	122	24.1	176	23	AAU70945	Cyteline intercellu
28	122	24.1	375	21	AAW74405	Neisseria meningit
29	122	22.9	813	21	AAW01674	FIS2 protein sequ
30	121.5	22.8	112	23	AAU77106	Frog transforming
31	121.5	22.8	112	23	AAU77107	Frog transforming
32	121.5	22.8	112	23	AAU77108	Frog transforming
33	121	22.7	392	20	AAW38681	Neisseria meningit
34	121	22.9	392	20	AAW38681	Neisseria meningit
35	120	23.7	139	23	AAU76687	Novel human secre
36	120	22.5	134	22	AAU30506	Novel human secre
37	120	23.7	177	20	AAV37952	Amino acid sequen
38	120	23.7	362	21	AAV77512	Plasmodium DBL gen
39	120	23.7	374	14	AAW33734	Human genetic sequ
40	120	22.5	411	21	AAV77914	Plasmodium DBL gen
41	120	22.5	443	20	AAV31670	Human IgG2 chain C
42	120	23.7	443	20	AAV31670	Human IgG2 chain C
43	120	22.5	444	20	AAV31672	Human IgG4 chain C
44	120	23.7	444	20	AAV31672	Human IgG4 chain C
45	120	22.5	447	20	AAV31669	Human IgG1 chain C

OY	138	CNCTGT	143
Dd	1	***Cys	2
 RESULT 2			
ID	AAU30663	standard; Protein:	158 AA.
XX	AAU30663		
AC	AAU30663:		
XX			
DT	18-DEC-2001	(first entry)	
XX			
DE	Novel human secreted protein #1154.		
XX			
KW	Human: vaccination; gene therapy; nutritional supplement;		
RW	stem cell proliferation; haematopoiesis; nerve tissue regeneration;		
KM	immune suppression; immune stimulation; anti-inflammatory; leukaemia.		
OS	Homo sapiens.		
XX			
PD	WO200179449-A2.		
XX			
PD	25-OCT-2001.		
XX			
PF	16-APR-2001; 2001WO-US08656.		
XX			
PR	18-APR-2000; 2000US-0552929.		
PR	26-JAN-2001; 2001US-0770160.		
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Tang YT, Liu C, Drmanac RT.		
DR	WPI: 2001-611725/70.		
XX			
PS	Nucleic acids encoding a range of human polypeptides, useful in genetic		
PT	vaccination, testing and therapy -		
XX			
PS	Claim 20: Page 325-326; 765pp: English.		
XX			
CC	The invention relates to novel human secreted polypeptides. The		
CC	polypeptides and antibodies to the polypeptides are useful for		
CC	determining the presence of or predisposition to a disease associated		
CC	with altered levels of polypeptide. The polypeptides are also useful for		
CC	identifying agents (agonists and antagonists) that bind to them. Cells		
CC	expressing the proteins are useful for identifying a therapeutic agent		
CC	for use in treatment of a pathology related to aberrant expression or		
CC	physiological interactions of the polypeptide. Vectors comprising		
CC	the nucleic acids encoding the polypeptides and cells genetically		
CC	engineered to express them are also useful for producing the proteins.		
CC	The proteins are useful in genetic vaccination, testing and		
CC	therapy, and can be used as nutritional supplements. They may be used to		
CC	increase stem cell proliferation; to regulate haematopoiesis; and in		
CC	bone, cartilage, tendon and/or nerve tissue growth or regeneration;		
CC	immune suppression and/or stimulation; as anti-inflammatory agents; and		
CC	in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid		
CC	sequences of novel human secreted proteins of the invention.		
XX			
SQ	Sequence	158 AA;	
 Alignment Scores:			
Pred. No.:	1.49e-07	Length:	158
Score:	131.00	Matches:	9
Percent Similarity:	56.25%	Conservative:	0
Best Local Similarity:	56.25%	Mismatches:	7
Query Match:	24.58%	Indels:	0
DB:	22	Gaps:	0
 US-09-880-711-328 (1-350) x AAU30663 (1-158)			
OY	26	CCCCAGTAGCCCTGAAGTGTGGGCCCGCTACCCTGATACCCACGCA	73
Dd	25	ProAnSerProAnSnglyGlnGlyProAlaSpSerSerProSerGly	40

RESULT 3
AAW13888
ID AAW13888 standard: protein; 730 AA.
XX
AC AAW13888;
DT 13-MAY-1997 (first entry)
XX
DE Generic formula #1 of ant toxin of the invention.
XX
KW Toxin; ant; Bacillus thuringiensis; hymenopteran pest; pharaoh ant;
XX biological control; Monomorium pharaonis; generic.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 2 /label= Ala, Thr
FT Misc-difference 3 /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 4 /label= Leu, Ile
FT Misc-difference 5 /label= Asn, Gln
FT Misc-difference 7 /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 10 /note= "any naturally occurring amino acid except Cys, or
FT absent"
FT Misc-difference 11 /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 12 /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 14 /label= Asn, Gln
FT Misc-difference 15 /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 16 /label= Leu, Ile
FT Misc-difference 17 /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 18..24 /note= "any naturally occurring amino acid except Cys, or
FT absent"
FT Misc-difference 25..35 /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 36 /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 37..38 /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 39 /note= "any naturally occurring amino acid except Cys, or
FT absent"
FT Misc-difference 40 /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 41 /label= Glu, Asp
FT Misc-difference 42..44 /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 45 /label= Lys, Arg
FT Misc-difference 46..48 /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 49 /label= Lys, Arg
FT Misc-difference 50..51 /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 52..56 /note= "any naturally occurring amino acid except Cys, or
FT absent"
FT Misc-difference 57 /note= "any naturally occurring amino acid except Cys"

FT Misc-difference 58 /label= Lys, Arg, Glu, Asp
FT Misc-difference 59..62 /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 63 /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 64..72 /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 73 /label= Leu, Ile
FT Misc-difference 74..82 /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 83 /label= Leu, Ile
FT Misc-difference 84 /label= Gly, Ser
FT Misc-difference 85 /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 86 /label= Leu, Ile
FT Misc-difference 87 /label= Gly, Ser
FT Misc-difference 88 /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 89 /note= "any naturally occurring amino acid except Cys, or
FT absent"
FT Misc-difference 90 /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 92..102 /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 103..104 /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 105..106 /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 107 /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 108..118 /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 120 /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 121..122 /note= "any naturally occurring amino acid except Cys, or
FT absent"
FT Misc-difference 123 /label= Leu, Ile
FT Misc-difference 124 /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 125..126 /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 127 /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 128..130 /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 131 /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 132..133 /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 134..135 /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 136..138 /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 139 /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 140..147 /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 148 /label= Asn, Gln
FT Misc-difference 149..150 /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 151 /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 152

FT Misc-difference 153 /note="any naturally occurring amino acid except Cys"
FT /label= Gly, Ser
FT Misc-difference 154 /label= Leu, Ile
FT Misc-difference 155 /label= Asn, Gln
FT Misc-difference 156.163 /note="any naturally occurring amino acid except Cys"
FT Misc-difference 164 /label= Ala, Thr
FT Misc-difference 165 /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 166.169 /note="any naturally occurring amino acid except Cys"
FT Misc-difference 170 /note="any naturally occurring amino acid"
FT Misc-difference 171.174 /note="any naturally occurring amino acid"
FT Misc-difference 175.189 /note="any naturally occurring amino acid except Cys, or
FT /note="absent"
FT Misc-difference 190.192 /note="any naturally occurring amino acid except Cys"
FT Misc-difference 193 /note="any naturally occurring amino acid"
FT Misc-difference 194.205 /note="any naturally occurring amino acid"
FT /note="any naturally occurring amino acid except Cys, or
FT /note="absent"
FT Misc-difference 206.207 /note="any naturally occurring amino acid except Cys"
FT /label= Asn, Gln
FT Misc-difference 208 /label= Gly, Ser
FT Misc-difference 209 /note="any naturally occurring amino acid except Cys"
FT /label= Ala, Thr
FT Misc-difference 213.214 /note="any naturally occurring amino acid except Cys"
FT /label= Leu, Ile
FT Misc-difference 216.217 /note="any naturally occurring amino acid except Cys"
FT /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 218 /note="any naturally occurring amino acid except Cys"
FT /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 219.222 /note="any naturally occurring amino acid except Cys, or
FT /note="absent"
FT Misc-difference 223 /note="any naturally occurring amino acid except Cys"
FT /label= Glu, Asp
FT Misc-difference 224 /label= Glu, Asp
FT Misc-difference 225.229 /note="any naturally occurring amino acid except Cys"
FT /label= Glu, Asp
FT Misc-difference 230.238 /note="any naturally occurring amino acid except Cys"

Alignment Scores:

Pred. No.: 1.37e-07 Length: 730
Score: 131.00 Matches: 2
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 24.588 Indels: 0
DB: 18 Gaps: 0

US-09-880-711-328 (1-350) x AAW13888 (1-730)

OY 24 TACCCC 29

Db 8 Typro 9

RESULT 4

AAW13888
ID AAW13888 standard; protein; 730 AA.
XX
AC AAW13888;
XX
DE 13-MAY-1997 (first entry)
XX
DE Generic formula #1 of ant toxin of the invention.
XX
KW Toxin; ant; Bacillus thuringiensis; hymenopterian pest; pharaoh ant;
KM biological control; Monomorium pharaonis; generic.
XX
OS Synthetic.
XX
EH Key
EH Location/Qualifiers
FT Misc-difference 2 /label= Ala, Thr
FT Misc-difference 3 /note="any naturally occurring amino acid except Cys"
FT Misc-difference 4 /label= Leu, Ile
FT Misc-difference 5 /label= Asn, Gln
FT Misc-difference 7 /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 10 /note="any naturally occurring amino acid except Cys, or
FT /note="absent"
FT Misc-difference 11 /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 12 /note="any naturally occurring amino acid except Cys"
FT Misc-difference 14 /label= Asn, Gln
FT Misc-difference 15 /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 16 /label= Leu, Ile
FT Misc-difference 17 /note="any naturally occurring amino acid except Cys"
FT Misc-difference 18.24 /note="any naturally occurring amino acid except Cys, or
FT /note="absent"
FT Misc-difference 25.35 /note="any naturally occurring amino acid except Cys"
FT Misc-difference 36 /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 37.38 /note="any naturally occurring amino acid except Cys"
FT Misc-difference 39 /note="any naturally occurring amino acid except Cys, or
FT /note="absent"
FT Misc-difference 40 /note="any naturally occurring amino acid except Cys"
FT /label= Glu, Asp
FT Misc-difference 41 /label= Glu, Asp
FT Misc-difference 42.44 /note="any naturally occurring amino acid except Cys"
FT Misc-difference 45 /label= Lys, Arg
FT Misc-difference 46.48 /note="any naturally occurring amino acid except Cys"
FT Misc-difference 49 /label= Lys, Arg
FT Misc-difference 50.51 /note="any naturally occurring amino acid except Cys"
FT Misc-difference 52.56 /note="any naturally occurring amino acid except Cys, or
FT /note="absent"
FT Misc-difference 57 /note="any naturally occurring amino acid except Cys"
FT /label= Lys, Arg, Glu, Asp

FT Misc-difference 59..62
FT /note= "any naturally occurring amino acid except Cys"
PT Misc-difference 63
PT /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 64..72
FT /note= "any naturally occurring amino acid except Cys"
PT Misc-difference 73
PT /label= Leu, Ile
FT Misc-difference 74..82
FT /note= "any naturally occurring amino acid except Cys"
PT Misc-difference 83
PT /label= Leu, Ile
FT Misc-difference 84
PT /label= Gly, Ser
FT Misc-difference 85
PT /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 86
PT /label= Leu, Ile
FT Misc-difference 87
PT /label= Gly, Ser
FT Misc-difference 88
PT /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 89
PT /note= "any naturally occurring amino acid except Cys, or absent"
FT Misc-difference 90
PT /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 92..102
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 103..104
PT /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 105..106
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 107
PT /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 108..118
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 120
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 121..122
PT /note= "any naturally occurring amino acid except Cys, or absent"
FT Misc-difference 123
PT /label= Leu, Ile
FT Misc-difference 124
PT /label= Met, Leu, Ile, Val, Phe
PT Misc-difference 125..126
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 127
PT /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 128..130
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 131
PT /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 132..133
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 134..135
PT /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 136..138
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 139
PT /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 140..147
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 148
PT /label= Asn, Gln
FT Misc-difference 149..150
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 151
PT /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 152
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 153

FT /label= Gly, Ser
FT Misc-difference 154
PT /label= Leu, Ile
PT Misc-difference 155
PT /label= Asn, Gln
FT Misc-difference 156..163
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 164
PT /label= Ala, Thr
FT Misc-difference 165
PT /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 166..169
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 170
PT /note= "any naturally occurring amino acid"
FT Misc-difference 171..174
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 175..189
PT /note= "any naturally occurring amino acid except Cys, or absent"
FT Misc-difference 190..192
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 193
PT /note= "any naturally occurring amino acid"
FT Misc-difference 194..205
PT /note= "any naturally occurring amino acid except Cys, or absent"
FT Misc-difference 206..207
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 208
PT /label= Asn, Gln
FT Misc-difference 209
PT /label= Gly, Ser
PT Misc-difference 210..211
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 212
PT /label= Ala, Thr
FT Misc-difference 213..214
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 215
PT /label= Leu, Ile
FT Misc-difference 215..217
PT /label= Met, Leu, Ile, Val, Phe
FT Misc-difference 218
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 219..222
PT /note= "any naturally occurring amino acid except Cys, or absent"
FT Misc-difference 223
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 224
PT /label= Glu, Asp
FT Misc-difference 225..229
PT /note= "any naturally occurring amino acid except Cys"
FT Misc-difference 230..238

Alignment Scores:

Pred. No.: 1.37e-07 Length: 730
Score: 131.00 Matches: 3
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 75.00% Mismatches: 1
Query Match: 25.89% Indels: 0
DB: 18 Gaps: 0

US-09-880-711-328 (1-350) x AAM13888 (1-730)

QY 217 NCATTCGCAAGC 206

Db 7 ***TyrPro*** 10

RESULT 5
ABG07680
ID ABG07680 standard; protein: 119 AA.

RESULT	6
ID	AAM13871
AC	AAM13871 standard; protein; 731 AA.
XX	AAM13871:
DT	13-MAY-1997 (first entry)
DE	Generic formula #2 of ant toxin of the invention.
XX	
KM	Toxin; ant; <i>Bacillus thuringiensis</i> ; hymenopteran pest; pharaoh ant;
OS	biological control; Monomorium pharaonis; generic.
XX	Synthetic.
FH	Key
FT	Misc-difference 2..7
FT	/note= "unspecified amino acids"
FT	Misc-difference 10..12
FT	/note= "unspecified amino acids"
FT	Misc-difference 14..90
FT	/note= "unspecified amino acids"
FT	Misc-difference 92..118
FT	/note= "unspecified amino acids"
FT	Misc-difference 120..229
FT	/note= "unspecified amino acids"
FT	Misc-difference 231..240
FT	/note= "unspecified amino acids"
FT	Misc-difference 242
FT	/note= "unspecified amino acid"
FT	Misc-difference 244..249
FT	/note= "unspecified amino acids"
FT	Misc-difference 251..319
FT	/note= "unspecified amino acids"
FT	Misc-difference 321..335
FT	/note= "unspecified amino acids"
FT	Misc-difference 336..342
FT	/note= "unspecified amino acids"
FT	Misc-difference 343..472
FT	/note= "unspecified amino acids"
FT	Misc-difference 474..541
FT	/note= "unspecified amino acids"
FT	Misc-difference 543..637
FT	/note= "unspecified amino acids"
FT	Misc-difference 638..644
FT	/note= "unspecified amino acids"
FT	Misc-difference 646..686
FT	/note= "unspecified amino acids"
FT	Misc-difference 688..728
FT	/note= "unspecified amino acids"
FT	Misc-difference 730..731
FT	/note= "unspecified amino acids"
PN	US5596071-A.
PD	21-JAN-1997.
PE	22-MAY-1991; 910S-0703977.

PR	24-NOV-1993.	9JUS-0158232.
PR	22-MAY-1991.	9JUS-0703977.
PR	25-NOV-1991.	9JUS-0797645.
PR	22-MAY-1992.	9JUS-0887980.
XX		
PA	(MYCO) MYCOGEN CORP.	
PI		
PI	Foncerrada L, Fu J, Kennedy MK, Meier H, Payne JM.	
PI	Randall JB, Schnepf HE, Schwab GE, Ulick HJ.	
XX		
XX		
DR	WPI, 1997-107615/10.	
XX		
CT	Bacillus thuringiensis toxin - active against hymenopteran pests	

PS Disclosure: Column 91-96: 64pp; English.
XX
CC AAW13888 and AAW13871 represent the generic formulae for the toxins of
the invention. The toxins of the invention are derived from the
CC *Bacillus thuringiensis* isolate PS8603 (NRRL B-18765), and are active
CC against hymenopteran pests. The toxins can be used for the biological
CC control of ants, particularly pharaoh ants (*Monomorium pharaonis*).
XX
SQ Sequence 731 AA:

Alignment Scores:
Pred. No.: 2.4e-07 Length: 731
Score: 129.00 Matches: 2
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 18.20% Indels: 0
DB: Gaps: 0

US-09-880-711-328 (1-350) x AAW13871 (1-731)

OY 24 TACCCG 29
|||||
DB 8 TyrPro 9

RESULT 7
AAW13871
ID AAW13871 standard; protein: 731 AA.
XX
AC AAW13871:
XX
DT 13-MAY-1997 (first entry)
XX
DE Generic formula #2 of ant toxin of the invention.
XX
KW Toxin: ant; *Bacillus thuringiensis*; hymenopteran pest; pharaoh ant;
KW biological control; *Monomorium pharaonis*; generic.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT 2..7
FT Misc-difference /note= "unspecified amino acids"
FT Misc-difference 10..12 /note= "unspecified amino acids"
FT Misc-difference 14..90 /note= "unspecified amino acids"
FT Misc-difference 92..118 /note= "unspecified amino acids"
FT Misc-difference 120..229 /note= "unspecified amino acids"
FT Misc-difference 231..240 /note= "unspecified amino acids"
FT Misc-difference 242 /note= "unspecified amino acids"
FT Misc-difference 244..249 /note= "unspecified amino acid"
FT Misc-difference 251..319 /note= "unspecified amino acids"
FT Misc-difference 321..335 /note= "unspecified amino acids"
FT Misc-difference 336..342 /note= "unspecified amino acids"
FT Misc-difference 343..472 /note= "unspecified amino acids"
FT Misc-difference 474..541 /note= "unspecified amino acids"
FT Misc-difference 543..637 /note= "unspecified amino acids"
FT Misc-difference 638..644 /note= "unspecified amino acids"
FT Misc-difference 646..686 /note= "unspecified amino acids"
FT Misc-difference 688..728 /note= "unspecified amino acids"

FT /note= "unspecified amino acids"
FT Misc-difference 730..731
FT /note= "unspecified amino acids"
XX
PN US5556071-A.
XX
PD 21-JAN-1997.
XX
PF 22-MAY-1991; 91US-0703977.
XX
PK 24-NOV-1993; 93US-0158232.
PK 22-MAY-1991; 91US-0703977.
PK 25-NOV-1991; 91US-0797645.
PK 22-MAY-1992; 92US-0887980.
PA (MYCO) MYCOGEN CORP.
XX
PI Poncerrada L, Pa J, Kennedy MK, Meier H, Payne JM;
PI Kandall JB, Schnepf HE, Schwab GE, Ulick RJ;
XX
DR WPT; 1997-107915/10.
XX
PT *Bacillus thuringiensis* toxin - active against hymenopteran pests
XX
PS Disclosure: Column 91-96: 64pp; English.
XX
CC AAW13888 and AAW13871 represent the generic formulae for the toxins of
the invention. The toxins of the invention are derived from the
CC *Bacillus thuringiensis* isolate PS8603 (NRRL B-18765), and are active
CC against hymenopteran pests. The toxins can be used for the biological
CC control of ants, particularly pharaoh ants (*Monomorium pharaonis*).
XX
SQ Sequence 731 AA:

Alignment Scores:
Pred. No.: 2.4e-07 Length: 731
Score: 129.00 Matches: 3
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 75.00% Mismatches: 0
Query Match: 25.49% Indels: 0
DB: Gaps: 0

US-09-880-711-328 (1-350) x AAW13871 (1-731)

OY 217 NCATCCGACNG 206
||||:|||||
DB 7 ***TyrPro*** 10

RESULT 8
AAW13907
ID AAW13907 standard; Protein: 731 AA.
XX
AC AAW13907:
XX
DT 13-NOV-2000 (first entry)
XX
DE Generic formula of formulaid toxins.
XX
KW hymenopteran; ant; pest control; 86Q3a; 17a; 17b; 33P2; 63B;
KW formulaid.
XX
OS unidentified.
XX
FH Key Location/Qualifiers
FT 1 Misc-difference 1..731 /note= "Xaa = any amino acid"
FT
XX
PN US6077937-A.
XX
PD 20-JUN-2000.
XX
PI 16-OCT-1998; 98US-0173891.
XX

PR 24-NOV-1993: 93US-0158232.
PR 22-MAY-1991: 91US-0703977.
PR 25-NOV-1991: 91US-0797645.
PR 22-MAY-1992: 92US-0887980.

XX
PA (MYCO) MYCOGEN CORP.
XX
XX
PI Meier H, Kennedy MK, Schwab GE, Fu J, Payne JM, Uick HJ;
PI Foncecerra L, Schnepf HE, Randall JB;
XX
WR: 2000-450960/39.

XX
DR New Bacillus thuringiensis toxins with activity against hymenopteran
PT pests such as fire ants and carpenter ants, conform to a specific
PT generic formula and have a specific amino acid sequence -
XX
XX

PS Claim 1; Column 93; 67pp; English.

CC The present invention relates to novel Bacillus thuringiensis toxins
CC with hymenopterian activity. Preparations containing protein from
CC Bacillus thuringiensis were tested for toxicity to ants. The N-termi
CC amino acids of toxic proteins were then sequenced. These sequences w
CC used to design oligonucleotide probes. The probes were used to clone
CC ant-active toxin genes. The toxic proteins can be used to control
CC pests such as fire ants, carpenter ants, Argentine ants and pharaoh
CC ants. The proteins can also be used for producing transgenic plants
CC that are resistant to attack by ants. The proteins are a safe and
CC effective biological control agent against ant pests. The present
CC sequence is the generic formula of the formicidal toxins.

XX
Sequence 731 AM;

Alignment Scores:	
Pred. No.:	2,4e-07
Score:	129.00
Percent Similarity:	100.00%
Best Local Similarity:	75.00%
Query Match:	25,49%
DB:	21
Length:	731
Matches:	3
Conservative:	1
Mismatches:	0
Indels:	0
Gaps:	0

US-09-880-711-328 (1-350) x AAB13907 (1-731)

```

QY      217 NCATTCACGNG 206
      |||::|||||
Db      7 ***TYRPro*** 10

```

RESULT 10
AAY77913

AC	AAV77913;
XX	
DT	13-JUN-2000 (first entry)
XX	

Plasmodium DBL gene family Proj3 conserved domain F1

KW DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
KW DABP: sialic acid binding protein; ebvna; EBV nuclear antigen 4; EBNA4

Plasmodium sp. OS

	Location/Qualifiers
FH Key	
Ft Misc-difference	1.411
Ft	/note= "residues indicated xaa are unspecified"

PN US5993827-A

PD 30-NOV-1999

PF 07-JUN-1995; 95US-0487826

PR 10-SEP-1993; 93US-0119677

XX (USSH) US EPEPT HEALTH & HUMAN SERVICES.
 PA
 XX
 P1 Sim KL, Chichnis C, Peterson DS, Su X, Wellens TE, Miller LH;
 XX
 DR WPI: 2000-154198/17.
 XX
 PT Isolated protein binding domains from Plasmodium vivax and Plasmodium
 PT falciparum erythrocyte binding proteins useful for vaccinating against
 PT malaria -
 XX
 PS Examples: Fig 1; 93pp; English.
 XX
 CC The invention relates to ebl-1 polypeptides that are encoded by the DBU
 CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially
 CC identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid
 CC Binding Protein (SABP), which are soluble proteins that appear in the
 CC culture supernatant after erythrocytes infected with malaria release
 CC merozoites. Immunochemical studies indicate that DABP and SABP are the
 CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
 CC and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
 CC used to vaccinate against malaria, especially caused by P. falciparum.
 CC Immunization with the polypeptide provides effective protection against
 CC malaria.
 CC
 SQ Sequence 411 AA:
 XX
 Alignment Scores
 Pred. No.: 3,266-07 Length: 411
 Score: 128.00 Matches: 5
 Percent Similarity: 88.89% Conservatave: 3
 Best Local Similarity: 55.56% Mismatches: 1
 Query Match: 25,30% Indels: 0
 DB: Gaps: 0
 XX
 US-09-880-711-328 (1-350) x AAY77913 (1-411)
 QY 148 TGTATACAGCCCTCTGGCATGCAGT 122
 ||||| ::|||::|
 DB 278 CYSIISSRCYSLNRYVAlaCysasn 286
 |||||
 RESULT 11
 ID AAO21480 standard: Protein: 440 AA.
 AC AAO21480;
 XX
 DT 15-AUG-2002 (first entry)
 DE
 DE Consensus protein sequence for NGR's.
 XX
 KW Cerebroprotective; neuroprotective; cytosstatic; Nogo receptor homologue;
 KW NGR3; axonal growth; central nervous system; CNS; cerebral injury;
 KW spinal cord injury; stroke; demyelinating disease; multiple sclerosis;
 KW monophasic demyelination; encephalomyelitis; Marchiafava-Bignami disease;
 KW multifocal leukoencephalopathy; panencephalitis; Spongy degeneration;
 KW Alexander's disease; Canavan's disease; metachromatic leukodystrophy;
 KW Krabbe's disease; immune; bait protein; genetic mapping; gene therapy;
 KW transgenic animal; unregulated cellular growth; cancer; tumour.
 KW
 XX
 OS unidentified.
 XX
 PN WO200229059-A2.
 XX
 PD 11-APR-2002.
 XX
 PP 06-OCT-2001: 2001WO-US31488.
 XX
 PP 06-OCT-2000: 2000US-238361P.
 XX
 PR (UYVA) UNIV YALE.
 PA (BIOJ) BIOGEN INC.
 PA
 XX

P1 StrLwattner SM, Cate RL, Sah DWY;
 XX
 DR WPI: 2002-415677/44.
 XX
 PT Novel Nogo receptor homology polypeptide, NGR2 or NGR3, useful for
 PT treating central nervous system disorder, cerebral injury, spinal cord
 PT injury, stroke, and demyelinating diseases -
 XX
 PS Example 16; Fig 1; 277pp; English.
 XX
 CC The invention relates to a Nogo receptor homologue polypeptide, NGR2 or
 CC NGR3, comprising a 50 amino acid LRRC2 sequence, a 284 amino acid NTLRRCT
 CC sequence, or a 420, 461 or 392 amino acid sequence, all given in the
 CC specification. The NGR3 protein or its binding antibody is useful for
 CC decreasing inhibition of axonal growth of a central nervous system (CNS)
 CC neuron, by contacting the neuron NGR3 or its antibody, and for treating
 CC CNS disease, disorder or injury. NGR3 or a vector comprising NGR3 is
 CC useful for treating cerebral injury, spinal cord injury, stroke,
 CC demyelinating diseases, e.g. multiple sclerosis, monophasic
 CC demyelination, encephalomyelitis, multifocal leukoencephalopathy,
 CC panencephalitis, Marchiafava-Bignami disease, Spongy degeneration,
 CC Alexander's disease, Canavan's disease, metachromatic leukodystrophy and
 CC Krabbe's disease. NGR3 is useful for inducing an immune response in a
 CC mammal against NGR3, as a bait protein in a two-hybrid or three-hybrid
 CC assay, and as a research tool for identification, characterisation and
 CC purification of interacting, regulatory proteins. The nucleotide
 CC sequences of the invention are useful for screening for RFP associated
 CC with certain disorders, for genetic mapping, and for gene therapy. The
 CC vector containing NGR3 is useful for producing non-human transgenic
 CC animals. The NGR3 binding antibody is useful for isolating and purifying
 CC NGR3, for localisation and/or quantitation of NGR3, and for diagnostic
 CC and therapeutic purposes. The sequences of the invention, vectors and
 CC antibodies are useful for treating or preventing unregulated cellular
 CC growth such as cancer and tumour growth. This sequence represents the
 CC consensus protein sequence for the NGR's of the invention.
 CC
 SQ Sequence 440 AA:
 XX
 Alignment Scores:
 Pred. No.: 4,36-07 Length: 440
 Score: 127.00 Matches: 3
 Percent Similarity: 100.00% Conservatave: 1
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 23.63% Indels: 0
 DB: Gaps: 0
 XX
 US-09-880-711-328 (1-350) x AAO21480 (1-440)
 QY 10 CAGTACCTACAC 21
 |||||
 DB 130 CUGTYFACCTYF 133
 |||||
 RESULT 12
 ID AAW37150 standard: Protein: 740 AA.
 AC AAW37150;
 XX
 DT 06-JUL-1998 (first entry)
 DE
 DE Murine Menu, Ene-Vasp like (Evl) protein consensus.
 XX
 KW Ene-VASP like protein; Evi protein; Menu; mammalian Ene; Enabled;
 KW cytoskeleton; cell morphology; cell adhesion; cell differentiation;
 KW cell growth; cell motility; mouse.
 KW
 XX
 OS Mus musculus.
 XX
 PN Key location/Qualifiers
 PA Domain 1..113
 PT /note= "Ene-VASP homology domain 1 (EVH1)"
 PT Misc-difference 8
 PT /note= "variable residue"

FT Misc-difference 12 /note= "variable residue"
FT Misc-difference 19..20 /note= "variable residues"
FT Misc-difference 26..29 /note= "variable residues"
FT Misc-difference 30..31 /note= "variable residues"
FT Misc-difference 37 /note= "variable residue"
FT Misc-difference 41..44 /note= "variable residues"
FT Misc-difference 56 /note= "variable residue"
FT Misc-difference 58 /note= "variable residue"
FT Misc-difference 65 /note= "variable residue"
FT Misc-difference 67 /note= "variable residue"
FT Misc-difference 94 /note= "variable residue"
FT Misc-difference 100..101 /note= "variable residues"
FT Misc-difference 104 /note= "variable residue"
FT Misc-difference 107..108 /note= "variable residues"
FT Misc-difference 112 /note= "variable residue"
FT Misc-difference 114..122 /note= "variable residues"
FT Misc-difference 124 /note= "variable residue"
FT Misc-difference 126..138 /note= "variable residues"
FT Misc-difference 140 /note= "variable residue"
FT Misc-difference 145 /note= "variable residue"
FT Misc-difference 148..150 /note= "variable residues"
FT Misc-difference 152 /note= "variable residue"
FT Misc-difference 155..239 /note= "variable residues"
FT Misc-difference 243 /note= "variable residue"
FT Misc-difference 245..413 /note= "variable residues"
FT Misc-difference 418..420 /note= "variable residues"
FT Misc-difference 422..423 /note= "variable residues"
FT Misc-difference 425..433 /note= "variable residues"
FT Misc-difference 438..444 /note= "variable residues"
FT Misc-difference 451..454 /note= "variable residues"
FT Misc-difference 460 /note= "variable residue"
FT Misc-difference 462..463 /note= "variable residues"
FT Misc-difference 466..468 /note= "variable residues"
FT Misc-difference 470..475 /note= "variable residues"
FT Misc-difference 477..478 /note= "variable residues"
FT Misc-difference 488..496 /note= "variable residues"
FT Misc-difference 498..513 /note= "variable residues"

FT Misc-difference 519 /note= "variable residues"
FT Misc-difference 528..530 /note= "variable residue"
FT Misc-difference 532 /note= "variable residues"
FT Misc-difference 535..562 /note= "variable residue"
FT Misc-difference 565 /note= "variable residues"
FT Misc-difference 572..574 /note= "variable residue"
FT Misc-difference 580..592 /note= "variable residues"
FT Misc-difference 594..601 /note= "variable residues"
FT Misc-difference 603..620 /note= "variable residues"
FT Misc-difference 622..623 /note= "variable residues"
FT Misc-difference 625 /note= "variable residue"
FT Misc-difference 629..630 /note= "variable residues"
FT Misc-difference 633..637 /note= "variable residues"
FT Misc-difference 639..641 /note= "variable residues"
FT Misc-difference 643..702 /note= "variable residues"
FT Misc-difference 707 /note= "variable residue"
FT Misc-difference 713 /note= "variable residue"
FT Misc-difference 715..717 /note= "variable residues"
FT Misc-difference 720 /note= "variable residue"
FT Misc-difference 731 /note= "variable residue"
FT Misc-difference 735..740 /note= "variable residues"
FT Misc-difference 735..740 /note= "variable residues"
FT Domain /note= "C-terminal homology domain"
PN W09801755-A1.
XX 15-JAN-1998.
XX 03-JUL-1997; 97WO-US11669.
XX 05-JUL-1996; 96US-0675815.
XX (GBFB) GBS BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX (HUTCHINSON) HUTCHINSON CANCER RES CENT FRED.
XX Gertler FB, Niebuhr K, Soriano P, Wehland J;
XX WPT, 1998-101197/09.
XX
XX Detection of modulators of Mena and Ena-VASP-like genes and proteins
XX - used in control of cytoskeletal dynamic events in normal and
XX abnormal cell morphology, adhesion, motility, growth and
XX differentiation
XX
XX Example 1: Page 68-70; 77pp; English.
XX
XX This polypeptide comprises a consensus of novel murine mammalian Ena
XX (Mena) protein (see AA037148) and novel murine Ena-VASP like protein
XX Evi (see AA037149). The greatest degree of amino acid identity is
XX in the N-terminal 'Ena-VASP homology domain' that is responsible
XX for Mena binding to Listeria, and to the cytoskeletal proteins

FT Misc-difference 370 amino acids"
 FT /label= Unknown
 FT /note= "X = any amino acid, residue represents Xm where
 FT m represents an integer"
 FT
 FT WO200157076-A1.
 FT
 FT PD 09-AUG-2001.
 FT
 FT PF 02-FEB-2001; 2001WO-NL00083.
 FT
 FT PR 04-FEB-2000; 2000GB-0002661.
 FT
 FT PA (NANO-) APPLIED NANOSYSTEMS BV.
 FT
 FT PI De Vocht ML, Woesten HAB, Wessels JGH;
 FT
 FT DR WPI; 2001-522347/57.
 FT
 FT PT purifying hydrophobin in hydrophobin containing solution by adsorbing
 FT the solution to surface, contacting the surface with a surfactant
 FT solution at specific temperature and separating desorbed hydrophobin
 FT from surface -
 FT
 FT PS Disclosure; Page 2; 10pp; English.
 FT
 FT CC The invention relates to purifying hydrophobin from a hydrophobin
 FT containing solution, comprising contacting the solution with a surface
 FT for adsorption and separating the adsorbed hydrophobin from the solution.
 FT CC Subsequently the surface is contacted with a solution containing a
 FT surfactant at a temperature lower than 900C. The present sequence is that
 FT of a generic hydrophobin protein sequence of the invention.
 FT
 FT SQ Sequence 370 AA:
 FT
 FT Alignment Scores:
 FT Pred. No.: 7.57e-07 Length: 370
 FT Score: 125.00 Matches: 3
 FT Percent Similarity: 100.00% Conservative: 0
 FT Best local Similarity: 100.00% Mismatches: 0
 FT Query Match: 24.70% Indels: 0
 FT DB: Gaps: 0
 FT
 FT US-09-880-711-328 (1-350) x AAG6541 (1-370)
 FT
 FT QY -190 TCCTGNAGN 182
 FT
 FT DB 2 Cys***** 4
 FT
 FT RESULT 15
 FT ID AAY79385
 FT AC AAY79385 standard; Protein; 176 AA.
 FT
 FT AC AAY79385;
 FT
 FT DT 01-AUG-2000 (first entry)
 FT
 FT XX EGF-like domain consensus.
 FT
 FT DE EGF-like domain consensus.
 FT
 FT XX SLGF; human G protein coupled receptor; signal transduction;
 FT antiapoptotic; cardiact; hypotensive; antiarteriosclerotic;
 FT antiarrhythmic; therapy; diagnosis; epidermal growth factor; EGF.
 FT
 FT OS Homo sapiens.
 FT
 FT FH Key Location/Qualifiers
 FT FT Misc-difference 1.4
 FT /note= "residues at positions 1-4 may be any amino
 FT acid"
 FT FT Misc-difference 5
 FT /note= "conserved Cys residue involved in disulf"
 FT FT Misc-difference 6.54

FT /note= "any 48 of the residues at positions 6-54
 FT may be absent; intended a range of 0-48
 FT amino acids"
 FT
 FT Disulfide-bond 55
 FT /note= "conserved Cys residue involved in a
 FT disulfide bond"
 FT
 FT Misc-difference 56.67
 FT /note= "any 9 of the residues at positions 56-67
 FT may be absent; intended a range of 3-12
 FT amino acids"
 FT
 FT Disulfide-bond 68
 FT /note= "conserved Cys residue involved in a
 FT disulfide bond"
 FT
 FT Misc-difference 69.138
 FT /note= "any 69 of the residues at positions 69-138
 FT may be absent; intended a range of 1-70
 FT amino acids"
 FT
 FT Disulfide-bond 139
 FT /note= "conserved Cys residue involved in a
 FT disulfide bond"
 FT
 FT Misc-difference 140.145
 FT /note= "any 5 of the residues at positions 140-145
 FT may be absent; intended a range of 1-6
 FT amino acids"
 FT
 FT Disulfide-bond 146
 FT /note= "conserved Cys residue involved in a
 FT disulfide bond"
 FT
 FT Misc-difference 151.171
 FT /note= "any 21 of the residues at positions 151-171
 FT may be absent; intended a range of 0-21
 FT amino acids"
 FT
 FT Misc-difference 173.174
 FT /note= "residues at positions 173-174 may be any
 FT amino acid"
 FT
 FT Disulfide-bond 175
 FT /note= "conserved Cys residue involved in a
 FT disulfide bond"
 FT
 FT Misc-difference 176
 FT /note= "residue at position 176 may be any amino
 FT acid"
 FT
 FT WO200018923-A2.
 FT
 FT PD 06-APR-2000.
 FT
 FT XX 30-SEP-1999; 99WO-US22923.
 FT
 FT PR 30-SEP-1998; 98US-0163821.
 FT
 FT PA (MILL-) MILLENNIUM PHARM INC.
 FT
 FT TSai F;
 FT
 FT WPI; 2000-293158/25.
 FT
 FT DR The present sequence is that of an epidermal growth factor (EGF)
 FT like domain consensus. In an embodiment of the invention, a novel
 FT human G protein coupled receptor, SLGF (see AAY79383), is identified
 FT on the basis of the presence of an EGF-like domain in the protein.
 FT The invention provides SLGF polynucleotides and polypeptides,
 FT antisense molecules, recombinant expression vectors, host cells,
 FT transgenic animals, fusion proteins, antigenic peptides,
 FT antibodies, and methods for identifying modulator compounds.
 FT Therapeutic and diagnostic methods utilizing these products are
 FT also provided, especially for disorders associated with aberrant
 FT SLGF expression such as proliferative, differentiative or

CC developmental, cardiac (hypertension, atherosclerosis coronary
CC artery disease, arrhythmias), and haematopoietic disorders.

XX
SQ Sequence 176 AA:

Alignment Scores:

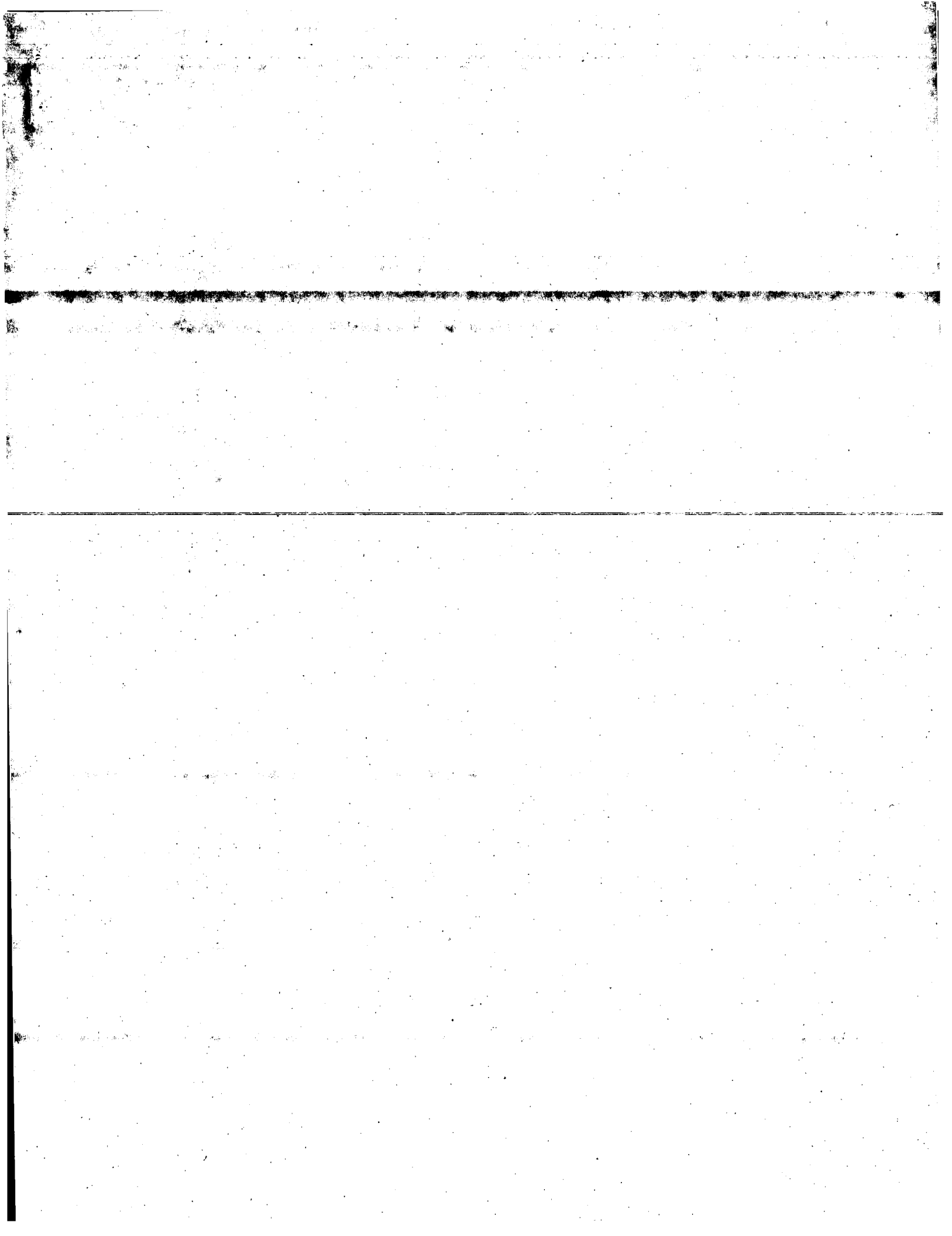
Pred. No.:	1.04e-06	Length:	176
Score:	124.00	Matches:	3
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	23.268	Indels:	0
Dh:	21	Gaps:	0

US-09-880-711-32E (1-350) x AAY79385 (1-176)

OY 254 GGAGCTTNN 262

Dh 149 G1YAla*** 151

Search completed: April 28, 2003, 15:36:04
Job time : 39 secs



GenCore version 5.1.4.F5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: April 28, 2003, 15:34:44 : Search time 15.5 Seconds
(without alignments)
3450.979 Million cell updates/sec

Title: JS-09-880-711-328

Perfect score: 533

Sequence: 1 atgagcagcagctacccgca.....ttgcgagatctcaataaaaa 350

Scoring table:

BLOSUM62DX	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5	
Fgapop 6.0	Fgapext 7.0	
Delop 6.0	Delext 7.0	

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xip
-O=/gen2_1/USPTC/US09860711/runat_28042003_112139_867/app_query.fasta_1.519
-DB=PIR_73 -QFMT=fastan -SUFIX=n2p.rpr -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blomsu62dx -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR.SCORE=PCL -THR.MAX=100 -THR.MIN=3 -ALIGN=15 -MODE=LOCAL
-OUTFMT=plc -NOM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09860711.acgn.1.1.62.gnumat.28042003_112139_867 -NCPUS=6 -ICPU=3
-NO_XIPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOC -DEV.TIMEOUT=120
-NARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: PIR_73:*
2: pir1:*
3: pir2:*
4: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	225.5	42.3	196	JC6127	RNA-binding protei
2	85.5	16.0	213	T13752	NADH2 dehydrogenas
3	79	15.6	213	T13752	NADH2 dehydrogenas
4	78.5	14.7	1870	S37671	MHC class III hist
5	78.5	14.7	1872	S36152	MHC class III hist
6	78.5	14.7	2142	B35098	MHC class III hist
7	75	14.1	212	T13571	NADH2 dehydrogenas
8	75	14.1	217	T13660	NADH2 dehydrogenas
9	74	13.9	228	S53504	unknown protein [1
10	74	13.7	652	T12587	extensin-like prot
11	73	13.5	579	E75275	NADH2 dehydrogenas
12	72	13.5	579	E75275	hypothetical prote
13	72	13.5	1159	T43461	probable phosphodi
14	70	13.1	494	S39607	transcription fact

15	70	13.1	542	2	S39608	transcription fact
16	70	13.8	578	2	T13664	NADH2 dehydrogenas
17	70	13.1	631	1	A36745	transcription fact
18	69.5	14.7	217	2	T13660	NADH2 dehydrogenas
19	69	12.9	427	2	G70555	hypothetical prote
20	68.5	12.9	427	2	E87669	conserved hypothet
21	68.5	12.9	1495	2	S60255	transcription co-r
22	68	12.8	428	2	T24769	hypothetical prote
23	68	14.4	652	2	T12587	NADH2 dehydrogenas
24	66.5	12.5	744	2	E86255	hypothetical prote
25	66.5	12.5	873	2	A47283	caliphilin - fruit
26	66	14.0	212	2	T13571	NADH2 dehydrogenas
27	66	12.4	407	2	C70816	hypothetical prote
28	66	12.4	578	2	T13664	NADH2 dehydrogenas
29	65.5	12.3	184	2	A57640	retinoblastoma bin
30	65	12.2	134	2	D84672	hypothetical prote
31	65	12.2	228	2	S46965	microfilament shea
32	65	12.2	356	1	WJH02H	homeotic protein H
33	65	12.2	519	2	T45764	hypothetical prote
34	65	12.8	544	2	S76859	hypothetical prote
35	64.5	12.7	184	2	T16044	hypothetical prote
36	64.5	12.7	232	2	JC7706	beta crystallin B1
37	64.5	12.1	395	2	H75457	hypothetical prote
38	64.5	12.1	751	2	S68957	adhesive plaque pr
39	64.5	12.1	1445	2	T50508	hypothetical prote
40	64	12.6	207	2	T12662	NADH2 dehydrogenas
41	64	12.0	212	2	T13367	microfilament shea
42	64	12.0	237	2	S46964	regulatory protein
43	64	12.0	432	2	A28913	hypothetical prote
44	64	12.0	447	2	T49350	hypothetical prote
45	63.5	11.9	233	2	T47136	hypothetical prote

ALIGNMENTS

RESULT 1
JC6127
RNA-binding protein type 1 - human
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: JC6127
R:Shimamoto, A.; Kikawa, S.; Ichikawa, K.; Suzuki, N.; Yamabe, Y.; Imanura, O.; Tokuta
Furuchi, Y.
Proc. Natl. Acad. Sci. U.S.A. 93, 10513-10517, 1996
A:Title: A unique human gene that spans over 230 kb in the human chromosome 8p11-12 a
A:Reference number: JC6127; MIMID:97008106; PMID:8855282
A:Accession: JC6127
A:Molecule type: mRNA
A:Residues: 1-196 <SH1>
A:Cross-references: DDBJ:D84107; NID:q1669546; PIDD:BA12225.1; PID:q1669547
C:Comment: This protein plays a role in RNA metabolism.
C:Genetics:
A:Gene: GDB:RBPMS
A:Cross-references: GDB:9456209; OMIM:601558
A:Map position: 8p12-8p11
A:Superfamily: human RNA-binding protein type 1; ribonucleoprotein repeat homology
C:Keywords: RNA binding
F:25-88/Domains: ribonucleoprotein repeat homology <RNM2>

Alignment Scores:
Pred. No.: 3,68e+18
Score: 225.50
Percent Similarity: 83.87%
Best Local Similarity: 76.15%
Query Match: 42.31%
DB: 1
US-09-880-711-328 (1-350) x JC6127 (1-196)

OY 3 GAGTCACAGTACCTGACCTTACCCAGTACGCTTGAAGTGTGGCCCGGACCTGTC 62
DB 134 GAGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT
134 GAGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT
153

Alignment Scores:	
Pred. No.:	1.09
Score:	75.00
Percent Similarity:	57.89%
Best Local Similarity:	19.30%
Query Match:	14.07%
DB:	2
Length:	217
Matches:	11
Conservative:	22
Mismatches:	10
Indels:	2
Gaps:	14

```

QY 178 NACINCTNCAGACATGATGATNGCGTCNCTGGGAATGNGCCCTGNCACAGATCAACTTA 237
      :::::::::::::::::::: |||::: |||::: |||
db 3 GlnAsnTySerSerThrLysLusSer-----LeuTySerLleSerLeu 18

```

QY 238 CTGCAGACCANCMAAGGAGCCTTNGACACAGATGTATTANNAACNACTACC-----291

Db 19 TTPGLYLSATGILEPROLYSELVALASNARGASPPHEALLLEUSERTHRTYLYSSER38

```

QY 292 -----TTATNCCTTTTGCAANGANAACINTT 31
      ::::: ::|||::|l|l|:::
Db 39 GlyValSerPhePheSerGlnAsnIleProLysIleGlnGly**Thra 55

```

```

RESULT 9
G86441
unknown protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
Accession: G86441

```

Rhede, L.G., 1990. *Journal of Polymer Science: Part A: Polymer Chemistry*, 28, 1567-1577.
 Rindfleisch, A.; Ecker, J.R.; Palm, C.J.; Federpspiel, N.A.; Kaul, S.; White, O.; Alonso, N.F.; Hughes, B.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.; Hiltz, L. 2000. *Journal of Polymer Science: Part A: Polymer Chemistry*, 38, 1816-1820.

C.A.: Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Matei, R.; Marcialli, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

A:Status: preliminary
A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis thaliana*
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: G86441

A:Residues: 1-1201 <STO>
 A:Cross-references: GB:AE005172; MID:g11136725; PIDN:AAG31306.L; GSPDB:GN001411
 C:Genetics:

[illegible]

US-09-880-711-328 (1-350) x G86441 (1-1201

```

15  CCAGCACTTTACCCAGTACCCCTGAAGTGTGGGCCCCGATCCCTCTGTACCCAGGAG 74
    ||| ||| ||||| |||
    ||| ||| ||||| |||
Db  702  ProProLeuProProSerSerThrArgLeuAlaProProProProProProPro 72

```

```
Db      722 LeuSerLysThrProAlaProProPro-ProLeuSerLysThrProValProProPr 74
```

```

Db      741 oProProlGlyLeuGIYArgGIYThrSerSergIYProProlGlyAlaLysGIyse 76
          |||         |||         ::
0v    1177 ANACTNCTNCAGCACATGATGATTCGGTCCGTGCGAATGCACGCCGACGATCAACCGC 73

```

Db 761 rasnalapProProProProProalaglyArggly-----77

Db 774 - - - - ArgAlaserLeuGlyIyArgGlyIyValIserValProThraIaIaIapr 79
Qy 297 CCTTTGCAANGANAACTNTTCCT 321

DB 792 Olyslstyrinalaleulysproleu 800
RESULT 10
S53504

C:Species: Medicago sativa (alfalfa)
C:Date: 15-Jul-1995 #sequence_revision 03-Nov-1995 #text_change 24-Nov-1999
C:Accession: S53504
P:Abraham S.; Hayes C.M.; Watson T.M.

Author: M. L. B. 2/1, 2/3-2/8, 1995
 A1: Title: Expression patterns of three genes in the stem of lucerne (Medicago sativa)
 A2: Reference number: S53504; MUID:95201245; PMID:7894016
 A3: Accession: S53504

A: Molecule type: mRNA
A: Residues: 1-228 <ABR>
A: Cross-references: GB:L36120; NID:g535585; PTDN:AB41815.1; PID:g53558

Alignment Scores:		
Pred. No.:	1.42	Length: 22
Score:	74.00	Matches: 26

```
Percent Similarity: 33.00%
Best Local Similarity: 24.53%
Query Match: 13.88%
DB: 2
Gaps: 3
```

US-09-880-711-328 (1-350) x S53504 (1-228)
QY 27 CCAGTAGCCCTGAAGTGTGGCCCCGTACC

0Y 81 CCGTCTTCCCTCCTCGCGCTTACCTACCCGCTTCACTGCATGCCACAGGCMC 14
 |||||
 00 121 PVIAlaserProProPhneserProProAlaIthrProProAlaIthrProPro 14
 |||||

141	TGTATACATCGCTAGNNNNNCAACCCCTG-----	ACNNNTGGAAACTNCTN	18
:	::::::::::	:	::::::::::

QY 186 CAGCACATGATGATNCGCTCMCTGGATGNGCCCTGCACAGATCAACTACTGCAGAC 24
::||| :::: |||::::: :::: :
db 179 uSsrpRoSerAsnAlaIproAlaproCyluEuSerCarIauSerProSerCylleuSerProCo 10

Db 998 PROProAlaPhaProProProthralaAsnIleMeProSerSerSerHisLeu-Phcgl 1017
QY 138 CMCCTGATACAGCCCTAGAGNNNNNACNCTGCACNNNTGGANCTCTCAGCAGCATGAT 197
Db 1017 YserMetProTProGlyProSerValProValProGlyLysProPhHisHisHisLeu 1037
QY 198 GATNGGCTGCTGCGATGCGCTGCACAGATCAACCTACTGACAGACANCAAGGAG 257
Db 1037 rSerGlyThrMetProMetAlaGlyGly-----IleProGlyGly 1050
QY 258 CTTNNGACAGACAGATGATTANNACCACTACC 291
Db 1050 yValHisAsnGlnPheIleProLeuGlnValThr 1061

RESULT 14

S39607
transcription factor HNF-1A, splice form C - human
C:Species: Homo sapiens (man)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 29-Aug-1997
C:Accession: S39607; S39752
R:Batch, I.; Yaniv, M.
EMBO J. 12, 4229-4242, 1993
A:Title: More potent transcriptional activators or a transdominant inhibitor of the HNF1
A:Reference number: S39607; MUID:94038905; PMID:7900999
A:Accession: S39607
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-494 <BAC>
A:Cross-references: EMBL:X71347
R:Batch, I.
submitted to the EMBL Data Library, April 1993
A:Reference number: S39752
A:Accession: S39752
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-494 <BAC>
A:Cross-references: EMBL:X71347
R:Batch, I.
submitted to the EMBL Data Library, April 1993
A:Reference number: S39752
A:Accession: S39752
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 118-494 <BA2>
A:Cross-references: EMBL:X71347
C:Genetics:
A:Gene: GDB:TCFL; HNF1; LFBI
A:Cross-references: GDB:125297; OMIM:142410; OMIM:600496
A:Map position: 12q24.3-12q24.3
C:Complex: homodimer; can also form heterodimers with, for example, HNF-1A form A
C:Function:
A:Description: transcription activator required for the expression of a number of liver
A>Note: this form is a more potent transcriptional activator than the A form
C:Superfamily: transcription factor HNF-1; homeobox homology
C:Keywords: alternative splicing; DNA binding; homeobox; liver; nucleus; transcription
F:1-33/Region: dimerization
F:222-277/Domain: homeobox homology <HOX>

Alignment Scores:
Pred. No.: 3.57 Length: 494
Score: 70.00 Matches: 29
Percent Similarity: 53.91% Conservative: 33
Best Local Similarity: 25.22% Mismatches: 38
Query Match: 13.13% Indels: 15
Gaps: 2

US-09-880-711-328 (1-350) x S39607 (1-494)

QY 24 TACCCAGTAGCCCTGAGAGTGTGGCCCGTACCTCTGTACCCAGCAGAGTTCAGCCCT 83
Db 286 TyrSerGlyProProProGlyProGlyProGlyProAlaLeuProAlaHisSerSerPro 305
QY 84 GCTCTTCCTCCTCCTCCGCTTCACCTACCCCGCTTCACTGCATGCCAGAGCCACTG 143
Db 306 GlyLeuProProProAla-LeuSerProSerLysValHisGlyValArgTyrGlyGln-- 324
QY 144 ATACATGCTAGAGNNNNNACNCTGCACNNNTGGANCTCTCAGCAGCATGATGATNG 203
Db 325 -----ProAlaThrSerGlnThrAlaGlnValProSerSerSerGlyGlyProLeuVal 342

QY 204 GTCNCTGGAGATNGCCTGNCACAGATC-----AACCT 236
Db 342 LThrValSerThrProLeuHisGlnValSerProThrGlyLeuGlnProSerHisSer 362
QY 237 ACTGCAGACCANCAAGG-----AGCTTNGACAGACAGATGATTANNACCACTAC 290
Db 362 uLeuSerThrGlnValAlaLysLeuValSerAlaAlaGlyGlyProLeuProProValSer 382
QY 291 CTTATNCCTTTTGCAANGANANACTTCTTGANATTTTG 333
Db 382 rLeuThrAlaLeuHisSerLeuGlnGlnThrSerProGlyLeu 396

RESULT 15

S39608
transcription factor HNF-1A, splice form B - human
N:Alternate names: hepatic nuclear factor 1-alpha; transcription factor 1, hepatic; t
C:Species: Homo sapiens (man)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 29-Aug-1997
C:Accession: S39608
R:Batch, I.; Yaniv, M.
EMBO J. 12, 4229-4242, 1993
A:Title: More potent transcriptional activators or a transdominant inhibitor of the H
A:Reference number: S39607; MUID:94038905; PMID:7900999
A:Accession: S39608
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-542 <BAC>
A:Cross-references: EMBL:X71346
C:Genetics:
A:Gene: GDB:TCFL; HNF1; LFBI
A:Cross-references: GDB:125297; OMIM:142410; OMIM:600496
A:Map position: 12q24.3-12q24.3
C:Complex: homodimer; can also form heterodimers with, for example, HNF-1A form A
C:Function:
A:Description: transcription activator required for the expression of a number of liv
A>Note: this form is a more potent transcriptional activator than the A form
C:Superfamily: transcription factor HNF-1; homeobox homology
C:Keywords: alternative splicing; DNA binding; homeobox; liver; nucleus; transcriptio
F:1-33/Region: dimerization
F:222-277/Domain: homeobox homology <HOX>

Alignment Scores:
Pred. No.: 3.51 Length: 542
Score: 70.00 Matches: 29
Percent Similarity: 53.91% Conservative: 33
Best Local Similarity: 25.22% Mismatches: 38
Query Match: 13.13% Indels: 15
Gaps: 3

US-09-880-711-328 (1-350) x S39608 (1-542)

QY 24 TACCCAGTAGCCCTGAGAGTGTGGCCCGTACCTCTGTACCCAGCAGAGTTCAGCCCT 83
Db 286 TyrSerGlyProProProGlyProGlyProGlyProAlaLeuProAlaHisSerSerPro 305
QY 84 GCTCTTCCTCCTCCTCCGCTTCACCTACCCCGCTTCACTGCATGCCAGAGCCACTG 143
Db 306 GlyLeuProProProAla-LeuSerProSerLysValHisGlyValArgTyrGlyGln-- 324
QY 144 ATACATGCTAGAGNNNNNACNCTGCACNNNTGGANCTCTCAGCAGCATGATGATNG 203
Db 325 -----ProAlaThrSerGlnThrAlaGlnValProSerSerSerGlyGlyProLeuVal 342
QY 204 GTCNCTGGAGATNGCCTGNCACAGATC-----AACCT 236
Db 342 LThrValSerThrProLeuHisGlnValSerProThrGlyLeuGlnProSerHisSer 362
QY 237 ACTGCAGACCANCAAGG-----AGCTTNGACAGACAGATGATTANNACCACTAC 290
Db 362 uLeuSerThrGlnValAlaLysLeuValSerAlaAlaGlyGlyProLeuProProValSer 382
QY 291 CTTATNCCTTTTGCAANGANANACTTCTTGANATTTTG 333
Db 382 rLeuThrAlaLeuHisSerLeuGlnGlnThrSerProGlyLeu 396

db 302 rleuthrAlaLeuHisSerLeuGlnGlnThrSerProGlyLeu 396

Search completed: April 28, 2003, 15:38:39
Job time : 23.5 secs

GenCore version 5.1.4.jp5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: April 28, 2003, 15:33:39 ; Search time 12 seconds
(without alignments)
2419.452 Million cell updates/sec

Title: US-09-880-711-328
Perfect score: 533
Sequence: 1 atgagctcacagctaccctga.....cttcagagatctcaataaadaa 350

Scoring table:
BLOSUM62DX
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-O=/cn2.1/USPnc_spool/US09880711/runat_28042003.112138.843/app_query.fasta.1.519
-DB=SwissProt_40 -OPMT=fasta -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOCPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62dx -TRANS=human40.cdi
-LIST=45 -DOCCALGN=200 -THR.SCORE=pcl -THR.MAX=100 -THR.MIN=0 -ALGN=15
-MODE=LOCAL -OUTFMT=pclo -NORM=ext -HEAPSIZE=500 -MHELEN=0 -MAXLEN=2000000000
-USER=US09880711 @CGC 1.1.26. @runat.28042003.112138.843 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAL -LARGEQUERY -NEG_SCORES=0 -WAIT -LONLOG -DEV_TIMECUT=120
-WARN_TIMECUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	240	45.0	197	1	RBMS_MOUSE
2	225.5	42.3	196	1	RBMS_HUMAN
3	151.5	28.4	196	1	RBMS_XENLA
4	146.5	27.5	200	1	RBMS_CHICK
5	119	23.5	551	1	IFEVA_HELPO
6	119	22.3	551	1	IFEVA_HELPO
7	119	22.3	2477	1	FINC_MOUSE
8	117	22.0	1877	1	PKCS_RAT
9	117	23.1	1877	1	PKCS_RAT
10	117	23.1	2477	1	FINC_MOUSE
11	115	21.6	265	1	CR2A_LYCES
12	115	21.6	265	1	CR2C_LYCES
13	115	21.6	267	1	CR2E_LYCES
14	115	21.6	267	1	CR2F_LYCES
15	114	21.4	591	1	GAG_MMTVC
16	107	21.1	265	1	CR2A_LYCES
17	107	21.1	265	1	CR2C_LYCES
18	107	21.1	267	1	CR2E_LYCES

C 19	107	21.1	267	1	CR2F_LYCES	P14277	lycopersico
C 20	104.5	20.7	591	1	GAG_MMTVC	P11284	mouse mamma
C 21	92	17.3	371	1	CYB_ERYTA	O48085	eryx talari
C 22	91	16.0	371	1	CYB_ERYTA	O48085	eryx talari
C 23	78.5	14.7	2142	1	BAT2_HUMAN	P48634	homo sapien
C 24	73	13.7	631	1	HNFA_HUMAN	P20823	homo sapien
C 25	71	14.3	371	1	CYB_ERYJA	O48076	eryx jaculu
C 26	70	14.8	371	1	CYB_ERYJA	O48076	eryx jaculu
C 27	68.5	12.9	2517	1	NSCR2_HUMAN	O95618	h nuclear r
C 28	68	13.4	720	1	PSA4_MARMO	O94mk1	marilia bo
C 29	66	14.0	233	1	SNCR4_MOUSE	O92112	mus musculu
C 30	65	12.2	356	1	HXH2_HUMAN	P14652	homo sapien
C 31	65	12.2	720	1	PSA4_MARMO	O94mk1	marilia bo
C 32	65	12.2	1394	1	CNG4_BOVIN	O28181	bos tauris
C 33	64.5	12.1	338	1	OTX3_BRARE	O90267	brachydanio
C 34	64.5	12.1	751	1	FPL_MYTGA	O27409	mytilus gal
C 35	64	12.0	432	1	BRLA_EMENI	P10069	emeticella
C 36	63.5	11.9	1163	1	RTN4_RAT	O91811	rattus norv
C 37	63	11.8	481	1	ETB2_HUMAN	O60883	homo sapien
C 38	63	11.8	512	1	GAG_SMSAV	P03330	simlan sarc
C 39	63	11.8	757	1	CIKE_HUMAN	O14003	homo sapien
C 40	63	11.8	1290	1	PER1_HUMAN	O15534	homo sapien
C 41	62.5	11.7	497	1	MAS2_HUMAN	O95645	homo sapien
C 42	62.5	11.7	591	1	MNT_MOUSE	O08789	mus musculu
C 43	62	11.6	87	1	CUO1_BACR	P80574	blaberus cr
C 44	62	11.6	520	1	RXR8_MOUSE	P28704	mus musculu
C 45	62	11.6	718	1	AT12_HSV11	P10230	herpes simp

ALIGNMENTS

RESULT 1
RBMS_MOUSE
ID RBMS_MOUSE STANDARD: PRT: 197 AA.
AC Q9WB0;
DT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE RNA-binding protein with multiple splicing (RBP-MS) (Heart, RRM
DE Expressed Sequence) (Hermes).
OS RBMS OR HERMES.
ON Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=99195799; PubMed=1096065;
RA Garber W.V., Yatskevych T.A., Antin P.B., Correia K.M., Conlon R.A.,
RA Krieger P.A.;
RT "The RNA-binding protein gene, hermes, is expressed at high levels in
the developing heart.";
RL Mech. Dev. 80:77-86(1999).
RL
CC -1- FUNCTION: MAY BIND RNA.
CC -1- TISSUE SPECIFICITY: RNA EXPRESSED IN DEVELOPING HEART, WITH
CC SIGNIFICANTLY HIGHER EXPRESSION IN THE ATRIA RELATIVE TO THE
CC VENTRICLES.
CC -1- DEVELOPMENTAL STAGE: RNA ALREADY DETECTED AT E9.5.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF148511; AAD39515.1; -.
CC MD5: MG1:1334446; Bdpms.
CC InterPro: IPR000504; RNA_rec_mot.
CC Pfam: PF00076; rrm; 1.

```
DR EMBL; D84107; BAA12225.1; -  
DR EMBL; D84108; BAA12226.1; -  
DR EMBL; D84109; BAA12227.1; -  
DR EMBL; D84111; BAA12229.1; -  
DR MIM; 601558; -  
DR InterPro; IPR000504; RNA_rec_mot.  
DR Pfam; PF00076; rrm; 1  
DR SMART; SMO0360; RRM; 1  
DR PROSITE; PS50102; RRM; 1  
DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.  
RW Alternative splicing; RNA-binding.  
FT DOMAIN 24 101  
FT VARSPLIC 177 196  
FT VARSPLIC 177 196  
FT VARSPLIC 177 196  
FT VARSPLIC 133 196  
FT VARSPLIC 133 176  
SQ SEQUENCE 196 AA; 21802 MW; 05DD964E62F8FB8C CRC64;  
  
Alignment Scores:  
Pred. NO.:  
Score: 4.64e-19 Length: 196  
Percent Similarity: 225.50 Matches: 46  
Best Local Similarity: 83.87% Conservative: 6  
Query Match: 74.19% Mismatches: 9  
DB: 42.31% Indels: 1 Gaps: 1  
  
US-09-880-711-328 (1-350) x RBMS_HUMAN (1-196)  
  
QY 3 GAGCGACAGTACTGTCGACCTTACCACAGTAGCCCTGAAGTGCGGCCGTGCCTCTG 62  
Db 134 GlutathioneS-transferase gamma class II cytosolic isoform 153  
QY 63 TACCAGGAGGATGAGCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 122  
Db 154 Tyrosyl-protein-ligase alpha class I isoform 172  
QY 123 CTGATGGCCGACGAGCCGCTGATATCATGCGCTAGGNNNNMCANCNCTGACNNNTGGANA 162  
Db 173 Leucine-rich repeat domain protein 192  
QY 183 CTCNCG 188  
Db 193 Argin 194  
  
RESULT 3  
RBMS_XENLA STANDARD: PRT; 196 AA.  
ID RBMS_XENLA AC O9YGP5;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE RNA-binding protein with multiple splicing homolog (RBP-MS) (Heart,  
DE RRM Expressed Sequence) (Hermes).  
GN RBMS OR HERMES.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI_Taxid=8355;  
[1]  
SEQUENCE FROM N.A.  
RP TISSUE=Heart;  
RC MEDLINE=9915799; PubMed=10096065;  
RA Gebel W.V., Yatskevych T.A., Antin P.B., Correia K.M., Conlon R.A.,  
RA Krieg P.A.;  
RT "The RNA-binding protein gene, hermes, is expressed at high levels in  
the developing heart";  
TL Mech. Dev. 80:77-86(1999).
```

```

CC -1- FUNCTION: MAY BIND RNA.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN DEVELOPING HEART, PRONEPHROS,
CC RETINA AND EPIDIDYMS. IN ADULT, HIGH EXPRESSION IN HEART, MODERATE
CC IN KIDNEY, UNDETECTABLE IN LIVER, LUNG AND SKELETAL MUSCLE.
CC -1- DEVELOPMENTAL STAGE: mRNA FIRST DETECTED IN THE TAILBUD EMBRYO
CC (STAGE 26) IN THE PAIRED HEART PRIMORDIA AND IN THE CONDENSING
CC EPITHELIUM THAT WILL FORM THE PRONEPHROS; AT THE LATE TAILBUD
CC STAGE (STAGE 34) IN THE DEVELOPING RETINA AND EPIDIDYMS. AS
CC DEVELOPMENT PROCEEDS, DETECTED THROUGH THE ENTIRE LENGTH OF THE
CC HEART TUBE, IN THE MUSCULAR TISSUE OF THE OUTFLOW TRACT, AND IN
CC THE DUCT EPITHELIUM OF THE PRONEPHROS. DURING LATER DEVELOPMENT,
CC MENA FOUND IN ALL SUBREGIONS OF THE HEART, IN THE GLOMUS, TUBULES
CC AND DUCT OF THE PRONEPHROS, IN THE RETINAL GANGLION CELL LAYER
CC (GCL) AND IN THE EPIDIDYMS.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF107489; A016971.1; .
CC DR HSSP: P09012; 201A.
CC DR InterPro: IPR000504; RNA_rec_mot.
CC DR Pfam: PF00076; rrm: 1.
CC DR SMART: SM00360; RRM: 1.
CC DR PROSITE: PS0102; RRM: 1.
CC DR PROSITE: PS00030; RRM_RNP_1; FALSE NEG.
CC KW RNA-binding.
CC FT DOMAIN 20 57 RNA-BINDING (RRM).
CC FT DOMAIN 11 14 POLY-ASN.
CC FT DOMAIN 167 172 POLY-ALA.
CC SO SEQUENCE 196 AA; 21675 MW; 7964198C1122A303 CRC64;

Alignment Scores:
Pred. No.: 2,73e-10 Length: 196
Score: 151.50 Matches: 35
Percent Similarity: 66.67% Conservative: 9
Best Local Similarity: 53.03% Mismatches: 17
Query Match: 28.42% Indels: 5
DB: Gaps: 2

US-09-880-711-328 (1-350) x RBMS_XENLA (1-196)
QY 3 GAGCTCAGAGTACCTGACCTTACCCAGTACGCGCTGAAGTGTGGCCCGCTACGCTCTG 62
DB 130 AspherhtrglAAlaAlaenllaprolaserProglAlaTPrAlaProTyrProleu 149
QY 63 TACCCAGCGAGTTAGCGCGCTGCTGCTGCTGCTGCGGCTTACGATACGCC----- 116
DB 150 TyrThrAlaGlnleAlaAlaProAlaIle--ProHisAlaAlaIaPheThrTyrProAlaIa 168
QY 117 -----GCTTCACTGCATGCCAGAGCGCTGTATACATGCTAGGNNNNACAGNCCTGAC 170
DB 169 AlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 188
QY 171 NNNTGGANACTNCTNCAG 188
DB 189 GlyTrrPlysserArgGln 194

RESULT 4
ID RBMS_CHICK STANDARD; PRT; 200 AA.
AC Q9W611;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE RNA-binding protein with multiple splicing homolog (RBP-MS) (Hfrt,
DE RRM Expressed Sequence) (Hermes).
GN RBMS OR HERMES.

```

```

OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=99195799; PubMed=10096065;
RA Gerdner W.V., Yatskevych T.A., Antin P.B., Correia K.M., Conlon K.A.,
RA Krieb P.A.;
RT "The RNA-binding protein gene, hermes, is expressed at high levels in
RT the developing heart.";
RL Mech. Dev. 80:77-86(1999).
CC -1- FUNCTION: MAY BIND RNA.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN DEVELOPING HEART.
CC -1- DEVELOPMENTAL STAGE: mRNA ALREADY DETECTED AT STAGE 7-8 IN THE
CC CARDIOGENIC MESODERM, AND BECOME ALMOST UNDETECTABLE IN THE OUTER
CC CIRCUMFERENCE OF THE VENTRICULAR REGION WHEREAS REMAINING HIGH IN THE
CC DEVELOPING ATRIAL REGIONS.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF129533; A030273.1; .
CC DR InterPro: IPR000504; RNA_rec_mot.
CC DR Pfam: PF00076; rrm: 1.
CC DR SMART: SM00360; RRM: 1.
CC DR PROSITE: PS0102; RRM: 1.
CC DR PROSITE: PS00030; RRM_RNP_1; FALSE NEG.
CC KW RNA-binding.
CC FT DOMAIN 22 59 RNA-BINDING (RRM).
CC FT DOMAIN 169 176 POLY-ALA.
CC SO SEQUENCE 200 AA; 21856 MW; 409383C780B8C1DF CRC64;

Alignment Scores:
Pred. No.: 1.07e-09 Length: 200
Score: 146.50 Matches: 34
Percent Similarity: 63.24% Conservative: 9
Best Local Similarity: 50.00% Mismatches: 18
Query Match: 27.49% Indels: 7
DB: Gaps: 2

US-09-880-711-328 (1-350) x RBMS_CHICK (1-200)
QY 3 GAGCTCAGAGTACCTGACCTTACCCAGTACGCGCTGAAGTGTGGCCCGCTACGCTCTG 62
DB 132 AspherhtrglAAlaAlaenllaprolaserProglAlaTPrAlaProTyrProleu 151
QY 63 TACCCAGCGAGTTAGCGCGCTGCTGCTGCTGCTGCGGCTTACGATACGCC----- 116
DB 152 TyrThrAlaGlnleAlaAlaProAlaIle--ProHisAlaAlaIaPheThrTyrProAlaIa 170
QY 117 -----GCTTCACTGCATGCCAGAGCGCTGTATACATGCTAGGNNNNACAGNCCTGAC 164
DB 171 AlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 190
QY 165 CCGAGCANNRGGANACTNCTNCAG 188
DB 191 GlnGlnGlyTrrPlysserArgGln 198

RESULT 5
ID IFEA_HELPO STANDARD; PRT; 551 AA.
AC P16274;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)

```


RESULT 8		
PCK5_RAT		
ID	PCK5_RAT	STANDARD: PRT: 1877 AA.
AC	P41413; Q62914;	
DT	01-NOV-1995 (Rel. 32, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	Protein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)	
DE	(Protein convertase PC5) (Subtilisin/kexin-like protease PC5)	
DE	(Convertase PC5) (PC5) (Fragments).	
OS	PCSK5.	
OS	Rattus norvegicus (Rat).	
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	

CC Mammalia, Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Rattus.
 CC NCBI_TaxId=10116;
 CC (1)
 CC SEQUENCE FROM N.A. (SHORT ISOFORM), AND TISSUE-SPECIFICITY.
 CC PC
 CC TISSUE=adrenal gland;
 CC MEDLINE=93342056; PubMed=8341687;
 CC RX
 CC Lussion J., Vieau D., Hamelin J., Day R., Chretien M., Seidah N.G.;
 CC RA
 CC "CDNA structure of the mouse and rat subtilisin/kexin-like PC5: a
 CC RT
 CC candidate proprotein convertase expressed in endocrine and
 CC RT
 CC nonendocrine cells";
 CC RT
 CC Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).
 CC (12)
 CC (13)
 CC SEQUENCE OF 1676-1877 FROM N.A. (LONG ISOFORM).
 CC PC
 CC TISSUE=adrenal gland;
 CC RA
 CC De Bie I., Martinkiewicz M., Nakayama K., Lazure C., Seidah N.G.;
 CC RT
 CC Submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.
 CC (13)
 CC SUBCELLULAR LOCATION.
 CC RX
 CC MEDLINE=20214819; PubMed=10749928;
 CC RA
 CC Xiang Y., Molloy S.S., Thomas L., Thomas G.;
 CC RT
 CC "The PC6B cytoplasmic domain contains two acidic clusters that direct
 CC RT
 CC sorting to distinct trans-Golgi network/endosomal compartments.";
 CC RT
 CC Mol. Biol. Cell 11:1257-1273(2000).
 CC (14)
 CC DEVELOPMENTAL EXPRESSION.
 CC RX
 CC MEDLINE=97166043; PubMed=9013936;
 CC RA
 CC Zheng M., Seidah N.G., Pitaru J.E.;
 CC RT
 CC "The developmental expression in the rat CNS and peripheral tissues of
 CC RT
 CC processes PC5 and PACE4 mRNAs: comparison with other proprotein
 CC RT
 CC processing enzymes.";
 CC RT
 CC Dev. Biol. 181:268-283(1997).
 CC (1)
 CC FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTASE ACTIVITY
 CC WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE
 CC OF CLEAVAGE AT THE RX(K/R) CONSENSUS MOTIF. MAY BE RESPONSIBLE
 CC FOR THE MATURATION OF GASTROINTESTINAL PEPTIDES. MAY BE INVOLVED
 CC IN THE CELLULAR PROLIFERATION OF ADRENAL CORTEX VIA THE ACTIVATION
 CC OF GROWTH FACTORS.
 CC (1)
 CC CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
 CC PROPEPTINS BY CLEAVAGE OF ARG-XAA-YAA-ARG-1-ZAA BONDS, WHERE XAA
 CC CAN BE ANY AMINO ACID AND YAA IS ARG OR IYS.
 CC (1)
 CC SUBCELLULAR LOCATION: PC5A IS SECRETED THROUGH THE REGULATED
 CC SECRETORY PATHWAY. PC5B IS A TYPE I MEMBRANE PROTEIN LOCALIZED TO
 CC A PARANUCLEAR POST-GOLGI NETWORK COMPARTMENT IN COMMUNICATION WITH
 CC EARLY ENDOSOMES (BY SIMILARITY).
 CC (1)
 CC ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; PC5B/LONG (SHOWN HERE)
 CC AND PC5A/SHORT; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC (1)
 CC TISSUE SPECIFICITY: EXPRESSED IN THE INTESTINE, BRAIN, ADRENAL
 CC GLAND, ANTERIOR PITUITARY, THYROID, OVARIES, TESTIS AND LUNG.
 CC (1)
 CC HIGHEST LEVELS ARE FOUND IN THE GUT, DUODENUM, JEJUNUM AND ILEUM.
 CC EXPRESSION IS HIGHER IN FEMALE THAN IN MALE REPRODUCTIVE ORGANS.
 CC DEVELOPMENTAL STAGE: FIRST DETECTED AT E9 IN HIGHLY RESTRICTED
 CC REGIONS OF THE NEURAL TUBE, IN CAUDAL MOTOMERES, AND AT THE
 CC MATERNO-EMBRYONIC JUNCTION OF THE UTERUS. AT E10, RESTRICTED
 CC EXPRESSION IS DETECTED IN THE OPTIC AND OPTIC VESICLES, THE ROOF OF
 CC MIDBRAIN, AND TRUNK MOTOMERES. BY MIDGESTATION (E13-E16),
 CC MULTIPLE REGIONS INCLUDING HIPPOCAMPUS, THALAMUS, HYPOTHALAMUS,
 CC BRAIN STEM, AND SPINAL CORD. EXPRESSION IS ALSO DETECTED IN
 CC SEVERAL PERIPHERAL ORGAN SYSTEMS, INCLUDING GUT, LUNG, ADRENAL AND
 CC KIDNEY PRIMORDIA.
 CC (1)
 CC DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
 CC ASSISTING THE FOLDING OF THE ZMOGEN WITHIN THE ENDOPLASMIC
 CC RETICULUM.
 CC (1)
 CC DOMAIN: AC 1 AND AC 2 (CLUSTERS OF ACIDIC AMINO ACIDS) CONTAIN
 CC SORTING INFORMATION. AC 1 DIRECTS TGN LOCALIZATION AND INTERACTS
 CC WITH THE TGN SORTING PROTEIN PACS-1.
 CC (1)
 CC SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
 CC (1)
 CC SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.lsb-sib.ch/announces/> or send an email to license@lsb-sib.ch).

 CC EMBL: L14933; AAA99906.1; -;
 DR EMBL: U47014; AAA87888.1; -;
 DR PIR: B48225; B48225.
 DR HSSP: Q99405; IMPT.
 DR MEROPS: S08_076; -;
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR002884; P_domain.
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR Pfam: PF01483; P; PARTIAL.
 DR PRINTS: PRO0723; SUBTILISIN.
 DR ProDom: PD000717; P_domain; 1.
 DR SMART: SM00261; FU; 6.
 DR PROSITE: PS00136; SUBTILASE_ASP; 1.
 DR PROSITE: PS00137; SUBTILASE_HIS; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
 KW Cleavage on pair of basic residues; Alternative splicing; Repeat;
 KW Transmembrane; 1.
 FT SIGNAL 1 34 BY SIMILARITY.
 FT PROPEP 35 116 BY SIMILARITY.
 FT CHAIN 117 1877 PROTEIN CONVERTASE SUBTILISIN/KEXIN TYPE 5.
 FT DOMAIN 117 1768 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 1769 1789 POTENTIAL.
 FT DOMAIN 1790 1877 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 117 452 CATALYTIC.
 FT DOMAIN 638 1753 HOMO B.
 FT DOMAIN 1825 1844 CYS-RICH MOTIF (CRM) REGION.
 FT DOMAIN 1856 1877 AC 1.
 FT SITE 116 117 AC 2.
 FT SITE 521 523 CLEAVAGE (AUTO-) (BY SIMILARITY).
 FT ACT_SITE 173 173 CELL ATTACHMENT SITE (POTENTIAL).
 FT ACT_SITE 214 214 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 388 388 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 227 227 N-LINKED (GLCNAC; .) (POTENTIAL).
 FT CARBOHYD 383 383 N-LINKED (GLCNAC; .) (POTENTIAL).
 FT CARBOHYD 667 667 N-LINKED (GLCNAC; .) (POTENTIAL).
 FT CARBOHYD 754 754 N-LINKED (GLCNAC; .) (POTENTIAL).
 FT CARBOHYD 804 804 N-LINKED (GLCNAC; .) (POTENTIAL).
 FT CARBOHYD 854 854 N-LINKED (GLCNAC; .) (POTENTIAL).
 FT CARBOHYD 1710 1710 N-LINKED (GLCNAC; .) (POTENTIAL).
 FT CARBOHYD 1732 1732 N-LINKED (GLCNAC; .) (POTENTIAL).
 FT VARSPLIC 878 915 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX -> ATEESMAEGCFMIVKRNILCQKRVLDQLCCKTCTPGC (IN ISOFORM PC5A).
 FT VARSPLIC 916 1877 MISSING (IN ISOFORM PC5A).
 FT SEQUENCE 1877 AA: 207888 MW: 890955DC6053444A CRC64:

Alignment Scores:

Pred. No.: 3.06e-06 Length: 1877
 Score: 117.00 Matches: 21
 Percent Similarity: 38.028 Conservative: 25
 Best Local Similarity: 17.368 Mismatches: 29
 Query Match: 21.958 Indels: 46
 DB: 1 Caps: 5

US-09-880-711-328 (1-350) x PKC5_RAT (1-1877)

QY 89 TCCTCTCTCTCGCGCTTTCACCTACCCCGCTTACTGCATCCAGAGGCMCTGAT--- 145
 Db 719 ThrSerSerCys-----ValAlaIIncysProGluGlySerTyrIIn 732
 QY 146 -----ACATGCTCAGGNNNN 160
 Db 733 AspllelyslYsAsnIlleCysGllYlYscYsserGluaEncyslYslYstYhGlyIhhe 752
 QY 161 CACMCTGTCACNNNTGAGNACTNCTNCAACACATGATGATNCTGCTNCTGCAATGNCGC 220

Db 753 HtsAsncYstHnGlnCysLysclYglYleuSerIeGlnGlySerArGcYsserValThr 772
 QY 221 -----TGMCACGATCAACCTAC 238
 Db 773 CysGluaSplcIYglInPhePheSerclYhNhsAspCysGlnProCysHlsArG-----Phe 790
 QY 239 TGCAACACACNCAAGGAGCGCTTNNCAACAGAGATGTATTANNAAACCACTACTATTATTC 298
 Db 791 CysAlaThrCysAlaGlyAlaGlyAlaAspGlYcYslleAsncYstHnGlnGlyTyrVal 810
 QY 299 TTT-----TGCAANACANACATNTCTTTCANAAATTTTGA 334
 Db 811 MetGluGluGlyArgCysValcIInSerCysSerValSerTyrTyrIleuAspHlsSerIeu 830
 QY 335 GAG 337
 Db 831 Glu 831
 RESULT 9
 PKC5_RAT
 ID PK5_RAT STANDARD: PKT; 1877 AA.
 AC P41413; G62914;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)
 DE (Proprotein convertase PC5) (Subtilisin/kexin-like protease PC5)
 DE (Convertase PC5) (PC5) (PC5) (Fragments).
 GN PCSK5.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A. (SHORT ISOFORM), AND TISSUE-SPECIFICITY.
 RP TISSUE=Adrenal gland;
 RC MEDLINE=93342056; PubMed=8341687;
 RA Lusson J., Vleau D., Hamelin J., Day R., Chretien M., Seidah N.G.;
 RT "cDNA structure of the mouse and rat subtilisin/kexin-like PC5: a
 RT candidate proprotein convertase expressed in endocrine and
 RT nonendocrine cells.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).
 RN [2]
 RP SEQUENCE OF 1676-1877 FROM N.A. (LONG ISOFORM).
 RP TISSUE=Adrenal gland;
 RC De Bie T., Marchukiewicz M., Nakayama K., Lazure C., Seidah N.G.;
 RL submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=20214819; PubMed=10749928;
 RA Xiang Y., Molloy S.S., Thomas L., Thomas G.;
 RT "The PC5b cytoplasmic domain contains two acidic clusters that direct
 RT sorting to distinct trans-Golgi network/endosomal compartments.";
 RL Mol. Biol. Cell 11:1257-1273(2000).
 RN [4]
 RP DEVELOPMENTAL EXPRESSION.
 RX MEDLINE=97166043; PubMed=9013936;
 RA Zheng M., Seidah N.G., Plinier J.E.;
 RT "The developmental expression in the rat CNS and peripheral tissues of
 RT proteases PC5 and PACED mRNAs: comparison with other proprotein
 RT processing enzymes.";
 RL Dev. Biol. 181:268-283(1997).
 CC -!- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY
 CC WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE
 CC OF CLEAVAGE AT THE RX(R/K/R)R CONSENSUS MOTIF. MAY BE RESPONSIBLE
 CC FOR THE MATURATION OF GASTROINTESTINAL PEPTIDES. MAY BE INVOLVED
 CC IN THE CELLULAR PROLIFERATION OF ADRENAL CORTEX VIA THE ACTIVATION
 CC OF GROWTH FACTORS.
 CC -!- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
 CC PROPROTEINS BY CLEAVAGE OF ARG-XAA-YAA-ARG-I-ZAA BONDS, WHERE XAA
 CC CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
 CC -!- SUBCELLULAR LOCATION: PC5A IS SECRETED THROUGH THE REGULATED

```

FT CARBOHYD 383 383 (POTENTIAL)
FT CARBOHYD 667 667 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 754 754 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 804 804 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 854 854 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1710 1710 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1732 1732 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT VASAPPLIC 878 915 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
FT
FT VASAPPLIC 916 1877 (IN ISOFORM PC5A)
FT VASAPPLIC 1877 AA: 207888 MW: 890955DC60534444 CRC64: MISSING (IN ISOFORM PC5A)
SQ SEQUENCE

Alignment Scores:
Pred. No.: 3.06e-06 Length: 1877
Score: 117.00 Matches: 11
Percent Similarity: 63.41% Conservative: 15
Best Local Similarity: 26.83% Mismatches: 12
Query Match: 23.12% Indels: 3
Db: Gaps: 1

US-09-880-711-328 (1-350) x PCK5_RAT (1-1877)
OY 235 GGTGATCTGTGNCAGCGNCATTCGCCAGNAGCGCATGATCATGATGCTGAGNAGTNTC 176
Db 761 G1yeuSerLeuG1ngLy-----SerTgcySerValThrcySG1uaSpG1yGln 777
OY 175 CANNNGTCAGGNGTGNMNNNCCTTAGGCATGTATACAGAGCCTCTGGGCGCTCAGTAGAGCG 116
Db 778 PhePheSerTgLyH1SaSpCyGcG1nProCysH1SaTgPheCysAlaThrcySaLaG1yAla 797
OY 115 GGG 113
Db 798 Gly 798

RESULT 10
FINC_MOUSE STANDARD; PRT: 2477 AA.
ID FINC_MOUSE
AC P11276; Q61568; Q61569; Q61567; Q64233;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Fibronectin precursor (FN) (Fragments).
GN FNL.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC NCB1_TaxID=10090;
CX
RN RN SEQUENCE OF 1-28 FROM N.A.
RP RP TISSUE=Liver;
RC RC STRAIN=NMRI;
RX RX MEDLINE=94131313; PubMed=8299972;
RA RA Polly P., Nicholson R.C.;
RT RT "Sequence of the mouse fibronectin-encoding gene promoter region.";
RN RN Gene 137:353-354(1993).
RL RL [2]
RP RP SEQUENCE OF 562-834 FROM N.A.
RC RC STRAIN=NMRI;
RX RX MEDLINE=95403556; PubMed=7673336;
RA RA Taitis J.P., Weller A., Timpl R., Ekblom M., Ekblom P.;
RT RT "Regulation of mesenchymal extracellular matrix protein synthesis by
transferring growth factor-beta and glucocorticoids in tumor
stroma.";
RL RL [3]
RP RP J. Cell Sci. 108:2153-2162(1995).
RL RL [4]
RP RP SEQUENCE OF 899-2376 FROM N.A.
RC RC Gotski G., Aros W., Norton P.;
RL RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RX RX [4]
RP RP SEQUENCE OF 2375-2477 FROM N.A.
RC RC MEDLINE=88124987; PubMed=3124113;
RA RA Blatti S.P., Foster D.N., Ranganathan G., Moses H.L., Getz M.J.;
RT RT "Induction of fibronectin gene transcription and mRNA is a primary

```

RT response to growth-factor stimulation of AKR-2B cells.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:1119-1123(1988).
 RL (5).
 RP SEQUENCE OF 2375-2477 FROM N.A.
 RC TISSUE=Kidney;
 KX MEDLINE=93011702; Pubmed=1327855;
 KA Khandjian E.W., Salomon G., Leonard N., Tremblay S., Turler H.;
 RT "Fibronectin gene expression in proliferating, quiescent, and SV40-
 RT infected mouse kidney cells";
 RL Exp. Cell Res. 202:464-470(1992).
 RN [6].
 RP STRUCTURE BY NMR OF 1447-1530.
 RX MEDLINE=98202578; Pubmed=9533887;
 RA Copie V., Tomita Y., Akiyama S.K., Aota S., Yamada K.M., Venable R.M.,
 RA Pastor R.W., Krueger S., Torchia D.A.;
 RT "solution structure and dynamics of linked cell attachment modules of
 RT mouse fibronectin containing the RGD and synergy regions: comparison
 RT with the human fibronectin crystal structure.";
 RL J. Mol. Biol. 277:663-682(1998).
 CC -1- FUNCTION: FIBRONECTIN BIND CELL SURFACES AND VARIOUS COMPOUNDS
 CC INCLUDING COLLAGEN, FIBRIN, HEPARIN, DNA, AND ACTIN. FIBRONECTINS
 CC ARE INVOLVED IN CELL ADHESION, CELL MOTILITY, OPSONIZATION, WOUND
 CC HEALING, AND MAINTENANCE OF CELL SHAPE.
 CC -1- SUBUNIT: MOSTLY HETERODIMERS OR MULTIMERS OF ALTERNATIVELY SPLICED
 CC VARIANTS, CONNECTED BY 2 DISULFIDE BONDS NEAR THE CARBOXYL ENDS;
 CC TO A LESSER EXTENT HOMODIMERS.
 CC -1- ALTERNATIVE PRODUCTS: EACH OF THE "EXTRA DOMAIN" & "THE CONNECTING
 CC STRAND 3 ARE PRESENT IN SOME FORMS OF FIBRONECTIN AND ABSENT IN
 CC OTHERS. THESE DIFFERENCES ARE DUE TO ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: PLASMA FN (SOLUBLE DIMERIC FORM) IS SECRETED
 CC BY HEPATOCYTES. CELLULAR FN (DIMERIC OR CROSS-LINKED MULTIMERIC
 CC FORMS), MADE BY FIBROBLASTS, EPITHELIAL AND OTHER CELL TYPES, IS
 CC DEPOSITED AS FIBRILS IN THE EXTRACELLULAR MATRIX.
 CC -1- PTM: SULFATED (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE II DOMAINS.
 CC -1- SIMILARITY: CONTAINS 17 FIBRONECTIN TYPE III DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z23725; CAA80422.1; -;
 DR EMBL: X82402; CAA57796.1; -;
 DR EMBL: X93167; CAA63634.1; -;
 DR EMBL: M18194; AAA37636.1; -;
 DR EMBL: S45680; AAB23491.1; -;
 DR PIR: A31371; A31371;
 DR PIR: C60597; C60597;
 DR PDB: 1MPN; 29-APR-98.
 DR PDB: 2MPN; 29-APR-98.
 DR PDB: MGI; 95566; F01;
 DR InterPro: IFR000561; EGF-like.
 DR InterPro: IFR003961; FN_III.
 DR InterPro: IFR000562; FN_Type_II.
 DR InterPro: IFR000083; fibnctn1.
 DR InterPro: IFR003962; Pn111-repeat.
 DR Pfam: PF00049; fn1; 4.
 DR Pfam: PF00041; fn3; 17.
 DR PRINTS: PROCO14; FNTYPE11.
 DR SMART: SM00058; FN1; 4.
 DR SMART: SM00060; FN3; 14.
 DR PROSITE: PSC0022; EGF_1; 1.
 DR PROSITE: PSC0023; FIBRONECTIN_2; PARTIAL.
 DR PROSITE: PSC1253; FIBRONECTIN_1; 4.
 KM Glycoprotein; Plasma; Heparin-binding; Acute phase; Phosphorylation;
 KM Sulfation; Cell adhesion; Repeat; Alternative splicing; Signal;
 KM 3D-structure; 1 32 BY SIMILARITY.

FT	CHAIN	33	2477	FIBRONECTIN.
FT	DOMAIN	53	273	FIBRIN- AND HEPARIN-BINDING 1.
FT	DOMAIN	308	608	COLLAGEN-BINDING.
FT	DNA_BIND	506	1171	CELL-ATTACHMENT.
FT	DOMAIN	1357	1630	HEPARIN-BINDING 2.
FT	DOMAIN	1811	2081	FIBRIN-BINDING 2.
FT	DOMAIN	2256	2427	FIBRONECTIN TYPE-1 1.
FT	DOMAIN	51	96	FIBRONECTIN TYPE-1 2.
FT	DOMAIN	56	140	FIBRONECTIN TYPE-1 3.
FT	DOMAIN	140	185	FIBRONECTIN TYPE-1 4.
FT	DOMAIN	185	230	FIBRONECTIN TYPE-1 5.
FT	DOMAIN	230	272	FIBRONECTIN TYPE-1 6.
FT	DOMAIN	305	343	FIBRONECTIN TYPE-1 7.
FT	DOMAIN	345	404	FIBRONECTIN TYPE-1 8.
FT	DOMAIN	405	469	FIBRONECTIN TYPE-1 9.
FT	DOMAIN	468	516	FIBRONECTIN TYPE-1 10.
FT	DOMAIN	516	559	FIBRONECTIN TYPE-1 11.
FT	DOMAIN	559	602	FIBRONECTIN TYPE-1 12.
FT	DOMAIN	609	706	FIBRONECTIN TYPE-1 13.
FT	DOMAIN	707	808	FIBRONECTIN TYPE-1 14.
FT	DOMAIN	809	903	FIBRONECTIN TYPE-1 15.
FT	DOMAIN	904	994	FIBRONECTIN TYPE-1 16.
FT	DOMAIN	995	1084	FIBRONECTIN TYPE-1 17.
FT	DOMAIN	1085	1172	FIBRONECTIN TYPE-1 18.
FT	DOMAIN	1173	1264	FIBRONECTIN TYPE-1 19.
FT	DOMAIN	1265	1355	FIBRONECTIN TYPE-1 20.
FT	DOMAIN	1356	1446	FIBRONECTIN TYPE-1 21.
FT	DOMAIN	1447	1536	FIBRONECTIN TYPE-1 22.
FT	DOMAIN	1537	1630	FIBRONECTIN TYPE-1 23.
FT	DOMAIN	1631	1720	FIBRONECTIN TYPE-1 24.
FT	DOMAIN	1721	1810	FIBRONECTIN TYPE-1 25.
FT	DOMAIN	1811	1902	FIBRONECTIN TYPE-1 26.
FT	DOMAIN	1903	1993	FIBRONECTIN TYPE-1 27.
FT	DOMAIN	1992	2081	FIBRONECTIN TYPE-1 28.
FT	DOMAIN	2082	2201	CONNECTING STAND 3 (CS-3) (V REGION).
FT	DOMAIN	2202	2283	FIBRONECTIN TYPE-1 17.
FT	DOMAIN	2294	2338	FIBRONECTIN TYPE-1 10.
FT	DOMAIN	2339	2381	FIBRONECTIN TYPE-1 11.
FT	DOMAIN	2383	2426	FIBRONECTIN TYPE-1 12.
FT	SITE	1614	1616	CELL ATTACHMENT SITE.
FT	SITE	2181	2183	CELL ATTACHMENT SITE.
FT	DISULFID	561	589	BY SIMILARITY.
FT	DISULFID	587	599	BY SIMILARITY.
FT	DISULFID	2296	2325	BY SIMILARITY.
FT	DISULFID	2323	2335	BY SIMILARITY.
FT	DISULFID	2341	2368	BY SIMILARITY.
FT	DISULFID	2366	2378	BY SIMILARITY.
FT	DISULFID	2385	2405	BY SIMILARITY.
FT	DISULFID	2407	2423	BY SIMILARITY.
FT	DISULFID	2458	2458	INTERCHAIN (WITH 2458 OF OTHER CHAIN).
FT	DISULFID	2462	2462	INTERCHAIN (WITH 2462 OF OTHER CHAIN).
FT	CARBOHYD	1006	1006	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1243	1243	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1250	1250	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2198	2198	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	MOD_RES	2475	2475	PHOSPHORYLATION (BY SIMILARITY).
FT	CONFLICT	2440	2440	N->T (IN REF. 5).
SO	SEQUENCE	2477 AA;	271416 MW;	802H53H47AE2A2M4 CRR64;

Alignment Scores:

Pred. No.:	Length:	2477
Score:	117.00	Matches: 20
Percent Similarity:	48.11%	Conservative: 31
Best local Similarity:	18.87%	Mismatches: 35
Query Match:	23.12%	Indels: 20
DB:	1	Gaps: 3

US-09-880-711-328 (1-350) x F1NC_MOUSE (1-2477)

OY 342 GAAATTCCTCAAAANNTTCAGAAAGATTTCCTGCAAAAGCAGTATAGTACTGCG 283
 DB 1403 Glutuster11eserProSerAspAsnAlaValValIeuThrAsnIleuProGlyThr 1422

```

OY 282 TTNTATACATCTCTCNCNNAAGCT-----CCCTTNGTGT 244
DB 1423 GUTYRLEVALSERSEVALTYRGLNHLISGLUSERLLEPROLEUNGRLY 1442
OY 243 CTGCAGTAGGTGATCTGTGCAG-----GCCATTC 211
DB 1443 ARGGLNLYSTRHGLYLEASPERPROTHRGLYPHEASPSERASPILETHRALASN 1462
OY 210 CAGNACGNCNATCATCTGCTGNAGNAGTNTCCANNNGCAGNGTNNNCTAGG 151
DB 1463 SERPETHVALHISTIPVALALAPROARGALAPROILETHGLTYRILELARGHIS 1482
OY 150 CATGATACAGNGCCTGTGGCAGTCAAGCGGGTAGCTGAACCGCAGAGAGAG 91
DB 1483 HSS-----Alaglnhisservalglargproarglnasprargvalpro 1498
OY 90 GAAGAGAGCGCCCTACT 73
DB 1499 SERARGASNSERTLETHR 1504

RESULT 11
CB2A_LYCES STANDARD: PRT: 265 AA.
AC P14274;
DB 01-JAN-1990 (Rel. 13, Created)
DB 01-JAN-1990 (Rel. 13, Last sequence update)
DB 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chlorophyll A-B binding protein 1A, chloroplast precursor (LHCII type
DE I CAB-1A) (LHCP) (Fragments).
GN CAB1A.
OS Lycopersicon esculentum (tomato);
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. T6;
RX MEDLINE=86165866; PubMed=3007291;
RA Pichersky E., Bernatzky R., Tanksley S.D., Bredenhach R.B.,
RA Kausch A.P., Cashmore A.R.;
RT "Molecular characterization and genetic mapping of two clusters of
RT genes encoding chlorophyll a/b-binding proteins in Lycopersicon
RT esculentum (tomato).";
RL Gene 40:247-258(1985).
CC -1- FUNCTION: THE LIGHT-HARVESTING COMPLEX (LHC) FUNCTIONS AS A LIGHT
CC RECEPTOR. IT CAPTURES & DELIVERS EXCITATION ENERGY TO PHOTOSYSTEMS
CC WITH WHICH IT IS CLOSELY ASSOCIATED. THE N-TERMINUS OF THE PROTEIN
CC EXTENDS INTO THE STROMA WHERE IT IS INVOLVED WITH ADHESION OF
CC GRANAL MEMBRANES AND PHOTOREGULATED BY REVERSIBLE PHOSPHORYLATION
CC OF ITS THREONINE RESIDUES; BOTH ARE BELIEVED TO MEDIATE THE
CC DISTRIBUTION OF EXCITATION ENERGY BETWEEN PHOTOSYSTEMS I AND II.
CC -1- SUBUNIT: THE LHC COMPLEX CONSISTS OF CHLOROPHYLLS (A & B) AND
CC CHLOROPHYLL A-B BINDING PROTEINS.
CC -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M30615; AAA34149.1;
CC EMBL: M30616; AAA34150.1;
CC InterPro: IPR001344; Chlroa_Abbind.
CC Pfam: PF00504; Chlroa_B-bind. 1.
CC ProDom: PD000275; Chlroa_Abbind. 1.
CC ChoroPhyll; Photosynthesis; Photosystem I; Photosystem II;
CC Thylakoid; Membrane; Chloroplast; Transit peptide; Multigene-family;
CC Transmembrane; Phosphorylation.
CC TRANSIT 1 34 CHLOROPLAST.

```

```

FF CHAIN 35 265 CHLOROPHYLL A-B BINDING PROTEIN 1A.
FF TRANSMEM 151 171 POTENTIAL.
FF TRANSMEM 219 235
SQ SEQUENCE 265 AA; 28178 MW; 8F70B741DBD22A79 CRC64;

Alignment Scores:
Pred. No.: 5,71e-06 Length: 265
Score: 115.00 Matches: 7
Percent Similarity: 56.52% Conservative: 19
Best Local Similarity: 15.22% Mismatches: 20
Query Match: 21.58% Indels: 0
Gaps: 0

US-09-880-711-328 (1-350) x CB2A_LYCES (1-265)
OY 130 CCCAGAGCNCNCTGTACATGCTAGGNNNACACGCCGACNNTGANNCTNCACG 189
DB 196 PROLEUGLYLEUALAGLAPSPROGLUALAPHEALAGLULEULYVALLYSLULLEYS 215
OY 190 ACATGNATGATNCGCTGCGGATGNGCGTCGNCACAGATCAACCTACTGCAGAC 249
DB 216 ASNGLYARGLEUALAMECPHESERMETHPHEGLYPHEPHEVALGINALALEVALTHRG 235
OY 250 NAAGGAGCCTTNGAGAC 267
DB 236 LYSGLYPROLEUGLUALSN 241

RESULT 12
CB2C_LYCES STANDARD: PRT: 265 AA.
AC P14275;
DB 01-JAN-1990 (Rel. 13, Created)
DB 01-JAN-1990 (Rel. 13, Last sequence update)
DB 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chlorophyll A-B binding protein 1C, chloroplast precursor (LHCII type
DE I CAB-1C) (LHCP) (Fragments).
GN CAB1C.
OS Lycopersicon esculentum (tomato);
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. T6;
RX MEDLINE=86165866; PubMed=3007291;
RA Pichersky E., Bernatzky R., Tanksley S.D., Bredenhach R.B.,
RA Kausch A.P., Cashmore A.R.;
RT "Molecular characterization and genetic mapping of two clusters of
RT genes encoding chlorophyll a/b-binding proteins in Lycopersicon
RT esculentum (tomato).";
RL Gene 40:247-258(1985).
CC -1- FUNCTION: THE LIGHT-HARVESTING COMPLEX (LHC) FUNCTIONS AS A LIGHT
CC RECEPTOR. IT CAPTURES & DELIVERS EXCITATION ENERGY TO PHOTOSYSTEMS
CC WITH WHICH IT IS CLOSELY ASSOCIATED. THE N-TERMINUS OF THE PROTEIN
CC EXTENDS INTO THE STROMA WHERE IT IS INVOLVED WITH ADHESION OF
CC GRANAL MEMBRANES AND PHOTOREGULATED BY REVERSIBLE PHOSPHORYLATION
CC OF ITS THREONINE RESIDUES; BOTH ARE BELIEVED TO MEDIATE THE
CC DISTRIBUTION OF EXCITATION ENERGY BETWEEN PHOTOSYSTEMS I AND II.
CC -1- SUBUNIT: THE LHC COMPLEX CONSISTS OF CHLOROPHYLLS (A & B) AND
CC CHLOROPHYLL A-B BINDING PROTEINS.
CC -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M30617; AAA34151.1;
CC EMBL: M30618; AAA34152.1;

```

DR InterPro: IPR001344; Chloro_Abbind.
 DR Pfam: PF00564; chloro_b-bind: 1.
 DR ProDom: PD000275; Chloro_Abbind: 1.
 KW Chlorophyll: Photosynthesis; Photosystem I; Photosystem II;
 KW Thylakoid; Membrane; Chloroplast; Transit peptide; Multigene family;
 KW Transmembrane; Phosphorylation.
 FT TRANSIT 1 34 CHLOROPHYLL A-B BINDING PROTEIN 1C.
 FT CHAIN 35 265
 FT TRANSMEM 151 171 POTENTIAL.
 FT TRANSMEM 219 235
 SO SEQUENCE 265 AA: 28207 MW: D1255041DCA39BC9 CRC64:

Alignment Scores:
 Pred. No.: 5.71e-06 Length: 265
 Score: 115.00 Matches: 7
 Percent Similarity: 56.52% Conservative: 19
 Best Local Similarity: 15.22% Mismatches: 20
 Query Match: 21.58% Indels: 0
 DB: 1 Gaps: 0

US-09-880-711-328 (1-350) x CB2C_LYCES (1-265)

OY 130 CCGACAGCGCGTGTATACATGCTAGCGNNNNACACGCTGACNNNTGGANACTGTCGACG 189
 DB 156 ProlenGlyLeuAlaIaMetPheSerMetPheGlyPhePheValGlnAlaIleValThrcly 215
 OY 150 ACATGATGATGCGTCGCTGCGATGCGCTGCGACGATGACCTGACGACGACGACG 249
 DB 216 AsnGlyAraGLeuAlaIaMetPheSerMetPheGlyPhePheValGlnAlaIleValThrcly 235
 OY 250 NAAGGACGCTTNNAGAC 267
 DB 236 LysGlyProlenGluAsn 241

RESULT 13
 CB2E_LYCES
 ID CB2E_LYCES STANDARD: PRT: 267 AA.
 AC P14276:
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Chlorophyll A-B binding protein 3A, chloroplast precursor (LHCII type
 DE I CAB-3A) (LHCP) (Fragments).
 GN CAB3A.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 CC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 OC NCBI_TaxID=4081:
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. T6;
 RX MEDLINE=86165866; PubMed=3007291;
 RA Pichersky E., Bernatzky R., Tanksley S.D., Breidenbach R.B.,
 RA Kausch A.P., Cashmore A.R.:
 RT "Molecular characterization and genetic mapping of two clusters of
 RT genes encoding chlorophyll a/b-binding proteins in Lycopersicon
 RT esculentum (tomato)."
 RL Gene 40:247-258(1985).

-I- FUNCTION: THE LIGHT-HARVESTING COMPLEX (LHC) FUNCTIONS AS A LIGHT
 RECEPTOR, IT CAPTURES & DELIVERS EXCITATION ENERGY TO PHOTOSYSTEMS
 WITH WHICH IT IS CLOSELY ASSOCIATED. THE N-TERMINUS OF THE PROTEIN
 CC EXTENDS INTO THE STROMA WHERE IT IS INVOLVED WITH ADHESION OF
 CC GRANA MEMBRANES AND PHOTOREGULATED BY REVERSIBLE PHOSPHORYLATION
 CC OF ITS THREONINE RESIDUES. BOTH ARE BELIEVED TO MEDIATE THE
 CC DISTRIBUTION OF EXCITATION ENERGY BETWEEN PHOTOSYSTEMS I AND II.
 CC -I- SUBUNIT: THE LHC COMPLEX CONSISTS OF CHLOROPHYLLS (A & B) AND
 CC CHLOROPHYLL A-B BINDING PROTEINS.
 CC -I- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
 CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
 CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
 CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL: M30619; AAA34154.1; -;
 DR EMBL: M30620; AAA34155.1; -;
 DR InterPro: IPR001344; Chloro_Abbind.
 DR Pfam: PF00564; chloro_b-bind: 1.
 DR ProDom: PD000275; Chloro_Abbind: 1.
 KW Chlorophyll: Photosynthesis; Photosystem I; Photosystem II;
 KW Thylakoid; Membrane; Chloroplast; Transit peptide; Multigene family;
 KW Transmembrane; Phosphorylation.
 FT TRANSIT 1 34 CHLOROPHYLL A-B BINDING PROTEIN 3A.
 FT CHAIN 35 267
 FT TRANSMEM 153 173 POTENTIAL.
 FT TRANSMEM 221 237
 SO SEQUENCE 267 AA: 28468 MW: AHPD1BQCD87358B CRC64:

Alignment Scores:
 Pred. No.: 5.71e-06 Length: 267
 Score: 115.00 Matches: 7
 Percent Similarity: 56.52% Conservative: 19
 Best Local Similarity: 15.22% Mismatches: 20
 Query Match: 21.58% Indels: 0
 DB: 1 Gaps: 0

US-09-880-711-328 (1-350) x CB2E_LYCES (1-267)

OY 130 CCGACAGCGCGTGTATACATGCTAGCGNNNNACACGCTGACNNNTGGANACTGTCGACG 189
 DB 158 ProlenGlyLeuAlaIaMetPheSerMetPheGlyPhePheValGlnAlaIleValThrcly 217
 OY 150 ACATGATGATGCGTCGCTGCGATGCGCTGCGACGATGACCTGACGACGACGACG 249
 DB 218 AsnGlyAraGLeuAlaIaMetPheSerMetPheGlyPhePheValGlnAlaIleValThrcly 237
 OY 250 NAAGGACGCTTNNAGAC 267
 DB 236 LysGlyProlenGluAsn 243

RESULT 14
 CB2E_LYCES
 ID CB2E_LYCES STANDARD: PRT: 267 AA.
 AC P14277:
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Chlorophyll A-B binding protein 3B, chloroplast precursor (LHCII type
 DE I CAB-3B) (LHCP) (Fragments).
 GN CAB3B.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 CC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 OC NCBI_TaxID=4081:
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. T6;
 RX MEDLINE=86165866; PubMed=3007291;
 RA Pichersky E., Bernatzky R., Tanksley S.D., Breidenbach R.B.,
 RA Kausch A.P., Cashmore A.R.:
 RT "Molecular characterization and genetic mapping of two clusters of
 RT genes encoding chlorophyll a/b-binding proteins in Lycopersicon
 RT esculentum (tomato)."
 RL Gene 40:247-258(1985).

-I- FUNCTION: THE LIGHT-HARVESTING COMPLEX (LHC) FUNCTIONS AS A LIGHT
 RECEPTOR, IT CAPTURES & DELIVERS EXCITATION ENERGY TO PHOTOSYSTEMS
 WITH WHICH IT IS CLOSELY ASSOCIATED. THE N-TERMINUS OF THE PROTEIN
 CC EXTENDS INTO THE STROMA WHERE IT IS INVOLVED WITH ADHESION OF
 CC GRANA MEMBRANES AND PHOTOREGULATED BY REVERSIBLE PHOSPHORYLATION
 CC OF ITS THREONINE RESIDUES. BOTH ARE BELIEVED TO MEDIATE THE
 CC DISTRIBUTION OF EXCITATION ENERGY BETWEEN PHOTOSYSTEMS I AND II.

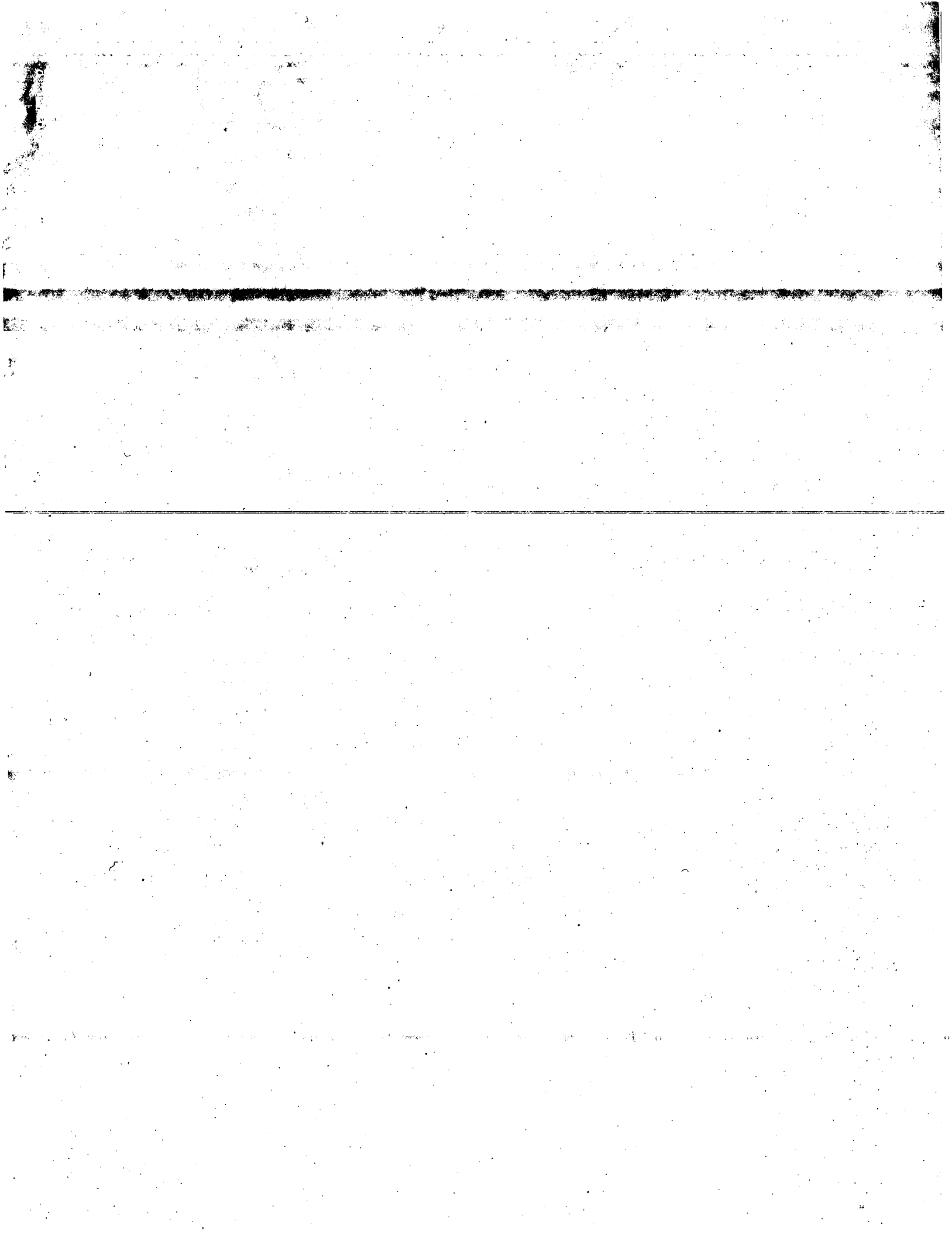
```

RA      Oroszian S.;
RT      "Analysis of gag proteins from mouse mammary tumor virus.";
RL      J. Virol. 63:2543-2549(1989).
CC       !- FUNCTION: P14 BINDS TO SINGLE STRANDED DNA.
CC       !- SIMILARITY: CONTAINS 2 CCHC-TYPE ZINC FINGERS.
CC       -----
CC       This SWISS-PROT entry is copyright. It is produced through a collabor
CC       between the Swiss Institute of Bioinformatics and the EMBL outpost
CC       at the European Bioinformatics Institute. There are no restrictions on
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for comme
CC       entities requires a license agreement (See http://www.isb.ch/anno
CC       or send an email to license@isb-sib.ch).
CC       -----
DR       EMBL; M16766; AAA66623.1; .
DR       PIR; A29029; A29029.
DR       InterPro: IPR003322; Gag_P10.
DR       InterPro: IPR000721; Gag_P24.
DR       InterPro: IPR001878; ZnF_CCHC.
DR       Pfam; PF00098; ZF-CCHC; 1.
DR       Pfam; PF00607; Gag_P24; 1.
DR       PRINTS; PF02337; Gag_P10; 1.
DR       PRINTS; PR00939; CZHCZNFINGER.
DR       SMART; SM00343; znf.C2HC; 2.
DR       PROSITE; PS50158; ZF_CCHC; 1.
KM       Coat protein; Core protein; Nucleoprotein; Polyprotein; Myristate;
FT       Phosphorylation; DNA-binding; Zinc-finger; Repeat.
FT       CHAIN          2
FT       CHAIN          99
FT       CHAIN          100   195    PHOSPHORYLATED PROTEIN PP21.
FT       CHAIN          196   228    PROTEIN P3.
FT       CHAIN          229   252    PROTEIN P8.
FT       CHAIN          270   496    MAJOR CORE PROTEIN P27.
FT       CHAIN          497   591    NUCLEIC ACID BINDING PROTEIN P14.
FT       ZN_FING        525   542    CCHC-TYPE 1.
FT       ZN_FING        552   569    CCHC-TYPE 2.
FT       LIPID          2
SQ       SEQUENCE      591 AA; 65338 MW; 4383324EDCC467C CRC64;
Alignment Scores:
Pred. No.:              7.26e-06           Length:         591
Score:                 114.00             Matches:         16
Percent Similarity:    50.50%             Conservative:    35
Best Local Similarity: 15.84%             Mismatches:     25
Query Match:          21.39%             Indels:         25
DB:                    1                   Gaps:           4

US-09-880-711-328 (1-350) x GAG_MMTVC (1-591)
QY      .84 GCCTTCCTCCTCCTCGGCCGCTTGACCTACC CGCGCTTGACTGCATGCCAGAGCG----- 138
            |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      404 AlaIlePProPrOfroGly -ValtVslYstThrValleuaLaclgLeuylngSlnglYaasgl 422
QY      139 -----NCTGTATAC-----ATGCCTAG 155
DB      423 uGlusertYrgluThrPhelleSerArgLeugLuGLuAlaValtyrryValMetProt 443
QY      156 G-----NNNNCAmCCTGACGANNNTGGANAACCTNCAGACCATGNATGAT 200
            |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      443 ggtlglygluYseraspIleleuIllelYsglLeuAlatrlplusaInlaasnsertleucY 463
QY      201 NCGCTGNCCTGGAAATNGCCCTGNCACAGATGAACCTACTGAGAGCCANCNAAGGACTT 266
            :|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      463 sGlnaspIleatrgPrometArgrlvYstHglylThmetGlnaspIrrileArGaIacy 483
QY      261 NNGAGAAGAAATGTATTNNNAACCACTACCTTTATNCTTTGGAAANGAANAACNTTCT 320
            ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      483 sleuasp-----AlaserProalavalValghlmYmetaIarYlaalaIaIame 500
QY      321 T 321
DB      500 t 500

```


Search completed: April 28, 2003, 15:36:36
Job time : 17 secs



GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame-plus.n2p model

Run on: April 28, 2003, 15:34:15 : Search time 36 Seconds

(without alignments)
4006.472 Million cell updates/sec

Title: US-09-880-711-328

Perfect score: 533

Sequence: 1 atgagcaccagctacccgca.....ctgagagattcacaataaaa 350

Scoring table:

BLOSUM62DX
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Zgapop 6.0, Zgapext 7.0
Delop 6.0, Delext 7.0

Searched: 571580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_n2p_model -DRV=xlp
-O=/cgn2_1/USPTO/US09880711/runat_28042003_11:138_855/app_query.fasta_1.519
-DB=SPTREMBL_21 -QPM=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloms62dx -TRANS=human40.cdi
-LIST=45 -DOCALLCN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NOR=ext -HEADSIZE=500 -MIFLEN=0 -MAXLEN=2000000000
-USER=US09880711.ecgn_1.1138.gunat_28042003_112136_855 -NCPU=6 -TCPU=3
-NO_XLPPX -NO_MKAP -LARGEQUERT -NEG_SCORES=0 -WAIT -LOGLOG -DRV=TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	240	45.0	197	11	Q921G0 mus musculus

2	240	45.0	220	11	Q9CPU5	Oryzops mus musculus
3	225.5	42.3	219	4	Q96J26	O96J27 homo sapien
4	125	24.5	478	8	Q8WJ67	Q8WJ67 tacca inteq
5	124	24.3	363	8	Q94X20	Q94X20 lysocochromi
6	124	24.5	431	8	P92380	P92380 galax urceo
7	124	24.5	442	8	Q8WJ66	Q8WJ66 burmannia c
8	124	24.5	512	8	Q9TKU2	Q9TKU2 andira gale
9	123	24.1	206	11	Q8VC52	Q8VC52 mus musculus
10	123	24.3	386	8	Q959C6	Q959C6 galaxias sp
11	123	24.3	478	8	Q8WJ67	Q8WJ67 lacca inteq
12	122	24.1	372	8	Q9XPW6	Q9XPW6 ceanochus o
13	122	22.9	409	8	Q9G5C9	Q9G5C9 melanoplus
14	122	22.5	460	8	Q8WJ65	Q8WJ65 thysania rod
15	122	24.1	505	8	Q19816	Q19816 allamanda c
16	122	24.1	505	8	Q8WJ21	Q8WJ21 rhyrachocaly
17	122	22.9	664	8	Q9BB53	Q9BB53 spatulatus
18	121	22.7	355	8	Q95FA6	Q95FA6 myodocarpus
19	121	23.9	363	8	Q94X20	Q94X20 lysocochromi
20	121	23.7	405	8	Q8SK37	Q8SK37 matillogobin
21	121	22.7	505	8	Q8WJ21	Q8WJ21 rhyrachocaly
22	121	22.7	507	8	Q8SK39	Q8SK39 pandaka iid
23	121	22.7	507	8	Q8SK38	Q8SK38 pandaka iid
24	120	23.7	340	8	Q03556	Q03556 stercorarii
25	120	23.7	355	8	Q95FA6	Q95FA6 myodocarpus
26	120	22.5	380	8	Q9G5C9	Q9G5C9 bellucla pe
27	120	22.5	439	8	Q33036	Q33036 rostkoviya c
28	120	22.5	442	8	Q8WJ65	Q8WJ65 burmannia c
29	120	23.7	446	8	Q92138	Q92138 distichia m
30	120	23.7	447	8	Q92K33	Q92K33 mastocarpus
31	120	23.7	447	8	Q9TKA3	Q9TKA3 pericalymna
32	120	22.5	508	8	Q8WJ20	Q8WJ20 dimorphanth
33	120	23.7	508	8	Q8WJ20	Q8WJ20 dimorphanth
34	119.5	22.4	515	8	Q9TK18	Q9TK18 dalbergia f
35	119	24.5	351	8	Q92832	Q92832 polyisipholi
36	119	22.3	372	8	Q9XPW6	Q9XPW6 ceanochus o
37	119	22.3	380	8	Q9G5C9	Q9G5C9 bellucla pe
38	119	22.3	431	8	P92380	P92380 galax urceo
39	119	22.3	436	8	Q32422	Q32422 hildebrand
40	119	22.3	447	8	Q32635	Q32635 mastocarpus
41	119	22.5	447	8	Q8WJ68	Q8WJ68 sphecelarita
42	119	22.5	449	8	Q35544	Q35544 kalliyenia
43	119	22.3	450	8	Q32803	Q32803 palcosia gla
44	119	22.3	450	8	Q32575	Q32575 metaspisope
45	119	23.5	457	8	Q8WJ22	Q8WJ22 pelagophycu

ALIGNMENTS

RESULT 1

ID	Q921G0	PRELIMINARY:	PRT:	197 AA.
AC	Q921G0:			
DT	01-DEC-2001 (TREMBLrel. 19, Created)			
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Unknown (protein for MGCI18922).			
GN	RBPM5.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Strasbourg R.			
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: BC011284; AAI11288.1; -			
DR	MGD: MGI:1334446; Rbpms-			
DR	InterPro: IPR00504; RNA_rec_mot.			
DR	Pfam: PF00076; rrm; 1.			
DR	PROSITE: PS50102; RRM; 1.			
SO	SEQUENCE 197 AA; 21816 MW; 03457084F803CC78 CRC64;			

Alignment Scores:

Pred. No.: 6,08e-22 Length: 197
Score: 240.00 Matches: 48
Percent Similarity: 85.48% Conservative: 5
Best Local Similarity: 77.42% Mismatches: 9
Query Match: 45.03% Indels: 0
DB: 11 Gaps: 0

US-09-880-711-328 (1-350) x Q921Q0 (1-197)

QY 3 GAGCTCAGTACCTGACCTTACCCAGTAGCCCTGAGTGTGGCCCGCTACCTCTG 62
|||||
Db 134 GluLeuThrValProAlaLeuTyrProSerSerProGluValTyrPalaProTyrProLeu 153
QY 63 TACCACGAGGAGTTAGCGCTGCTTCTCCTCCTGCGCTTACCTACCCGCTTCA 122
|||||
Db 154 TyrProAlaGluLeuAlaProAlaLeuProProProAlaAlaPheThrTyrProAlaSer 173
QY 123 CTGCATGCCAGAGGCGCTGTATACATGCTAGGNNNNACNCCCTGACNNNTGANA 182
|||||
Db 174 LeuHisAlaGlnMetArgTyrPheProProSerGluAlaThrSerGlnGlyTyrPlySer 193
QY 183 CTNCAAG 188
|||
Db 194 ArgGln 195

RESULT 2

Q9CPUS

ID Q9CPUS PRELIMINARY: PRT: 220 AA.

AC Q9CPUS:

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

GN 2700019M19R1K protein.

GN RBPMS OR 2700019M19R1K.

GN Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=EMBRIO;

RC MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Aizawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,

RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H. A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,

RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,

RA Schiml L. M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M. F.,

RA Brownstein M. J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RA Gustincich S., Hill D., Hofmann M., Hume D. A., Kamiya M., Lee N. H.,

RA Lyons P., Marchionni L., Mashima J., Mazzaroli J., Momberts P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K. F.,

RA Suzuki H., Toyooka K., Wang K. H., Whittaker C., Whittaker C., Wilming

RA Wyszawski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlski S.,

RA Hayashizaki Y.;

RA "Functional annotation of a full-length mouse cDNA collection."

RT Nature 409:685-690(2001).

RL EMBL: AK012586; BAB28336.1;

DR EMBL: AK012265; BAB28128.1;

DR MGD: MGI:133446; RBPMS.

DR InterPro: IPR000504; RNA_rec_mot.

DR Pfam: PF00076; rtm; 1.

DR SMART: SM00360; RRM; 1.

DR PROSITE: PS50102; RRM; 1.

SQ SEQUENCE 220 AA: 24291 MW: F5070707B142231BD CRC64:

Alignment Scores: 6.07e-22 Length: 220

Score: 240.00 Matches: 62
Percent Similarity: 87.36% Conservative: 14
Best Local Similarity: 71.26% Mismatches: 10
Query Match: 45.03% Indels: 1
DB: 11 Gaps: 1

US-09-880-711-328 (1-350) x Q9CPUS (1-220)

QY 3 GAGCTCAGTACCTGACCTTACCCAGTAGCCCTGAGTGTGGCCCGCTACCTCTG 62
|||||
Db 134 GluLeuThrValProAlaLeuTyrProSerSerProGluValTyrPalaProTyrProLeu 153
QY 63 TACCACGAGGAGTTAGCGCTGCTTCTCCTCCTGCGCTTACCTACCCGCTTCA 122
|||||
Db 154 TyrProAlaGluLeuAlaProAlaLeuProProProAlaAlaPheThrTyrProAlaSer 173
QY 123 CTGCATGCCAGAGGCGCTGTATACATGCTAGGNNNNACNCCCTGACNNNTGANA 182
|||||
Db 174 LeuHisAlaGlnMetArgTyrPheProProSerGluAlaThrSerGlnGlyTyrPlySer 193
QY 183 CTNCAAG 188
|||
Db 194 ArgGln 195

QY 243 GACCANCAAGGAGGAGCTTNG 263

Db 213 AspGlnGlnArgGluLeuPro 219

RESULT 3

Q96J26

ID Q96J26 PRELIMINARY: PRT: 219 AA.

AC Q96J26:

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE RNA-binding protein gene with multiple splicing.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=PLACENTA;

RA Strausberg R.;

RL Submitted (FE8-2001) to the EMBL/Genbank/DBJ databases.

DR EMBL: BC003608; AA03608.1;

DR InterPro: IPR000504; RNA_rec_mot.

DR Pfam: PF00076; rtm; 1.

DR PROSITE: PS50102; RRM; 1.

SQ SEQUENCE 219 AA: 24277 MW: 700E9B9FA7682FAD CRC64:

Pred. No.: 4.2e-20 Length: 219
Score: 225.50 Matches: 61
Percent Similarity: 86.21% Conservative: 14
Best Local Similarity: 70.11% Mismatches: 10
Query Match: 42.31% Indels: 2
DB: 4 Gaps: 2

US-09-880-711-328 (1-350) x Q96J26 (1-219)

QY 3 GAGCTCAGTACCTGACCTTACCCAGTAGCCCTGAGTGTGGCCCGCTACCTCTG 62
|||||
Db 134 GluLeuThrValProAlaLeuTyrProSerSerProGluValTyrPalaProTyrProLeu 153
QY 63 TACCACGAGGAGTTAGCGCTGCTTCTCCTCCTGCGCTTACCTACCCGCTTCA 122
|||||
Db 154 TyrProAlaGluLeuAlaProAlaLeuProProProAlaAlaPheThrTyrProAlaSer 172
QY 123 CTGCATGCCAGAGGCGCTGTATACATGCTAGGNNNNACNCCCTGACNNNTGANA 182
|||||
Db 173 LeuHisAlaGlnMetArgTyrPheProProSerGluAlaThrSerGlnGlyTyrPlySer 192

[illegible]


```

Db 97 Leuc1yarg1leheasValleuGlygluProvalAspasuLeuGlyProvalasp-Th 116
QY 225 TGNAGGCNCAATCC----- 210
Db 116 ratgtrThrPheProHisArgSerAlaProAlaPheIleGluLeuAspThrArgLe 136
QY 209 -----AGACAGCCNATCAATTCATGCTGCGTGNAGNAGTATCCANNNGTCAGGNGTGN 160
Db 136 User1lePheGluThrGlyIleLeuValValAspLeuAlaProTyArgArgGlyGly 156
QY 159 NNNCCTAGGCATGTATACAGCGCTTGCGGCATGCAGTACAGCGGGGTAGTGAAGC 102
Db 156 yls1leGlyLeuPheGlyGly-----AlaGlyValGlyGlyThr 169

RESULT 12
Q9XPM6 ID Q9XPM6 PRELIMINARY; PRT; 372 AA.
AC Q9XPM6;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Probable intron maturase (Maturase K) (Fragment).
GN MATK.
OS Ceanothus ophiocilius.
OC Chloroplast.
CC Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
CC eurosids I; Rosales; Rhamnaceae; Ceanothus.
OX NCBI_taxid=73700;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=158720PHIO;
RA Hardig T.M., Solitis P.S., Solitis D.E.;
RT "Diversification of the North American shrub genus Ceanothus
RT (Rhamnaceae): conflicting phylogenies from nuclear ribosomal DNA and
RT chloroplast DNA.";
RL Am. J. Bot. 0:0-0(1999).
CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS (BY SIMILARITY).
CC CC
CC -1- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
CC MITOCHONDRIAL INTRONS.
CC DR EMBL: AF049839; AAD30350.1; ".
CC DR InterPro: IPR002866; MatK_N.
CC Pfam: PF01824; MatK_N.1
CC KX Chloroplast; mRNA processing.
CC FT NON_TER 372 372
SQ SEQUENCE 372 AA; 43146 MW; DCCF7EBB963583A8 CRC64;

Alignment Scores:
Pred. No.: 5.69e-07 Length: 372
Score: 122.00 Matches: 14
Percent Similarity: 65.71% Conservative: 32
Best local Similarity: 20.00% Mismatches: 15
Query Match: 24.11% Indels: 9
DB: 8 Gaps: 3

US-09-880-711-328 (1-350) x Q9XPM6 (1-372)
QY 336 TCTCAANATTTNCAAGANAGTTCNTTTCACAAAGNATNAGTAGTNGTNTNTA 277
Db 91 SerGlnMetIleSerGluGlyPheAlaValIleValGluIleProPheSerLeuArgPhe 110
QY 276 ATACATTCTGCTCCNNAAGCTCCCTTGNTGNTGCTGCAGTAGGTGATCTGTGCNAGCGN 217
Db 111 LeuArgLeuValSerSerArgProGluIle--ValLysTyrSerAsnLeuArgSerIle 129
QY 216 CATTC-----GAGNAGCGNATCATNCAATGCTCTGTCNAGAGTNTCC 175
Db 130 HisSerIlePheProPheLeuGluAspLysPheProHisLeu-----AsnTyrGlySer 147
QY 174 ANNGTCAGGNGTGNNNNCCTAGCATGTA 145

```


Db 148 AsplleatgllleProtyrProtlehlsleu 157

RESULT 13

09G5C9 PRELIMINARY: PRT: 409 AA.

AC 09G5C9:

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1) (Fragment).

GN COI.

OS Melanoplus truchensis.

OC Mitochondrion.

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caellifera;

OC Acridomorpha; Acridoidea; Acrididae; Melanoplineae; Melanoplus.

OX NCBI_TaxID=12140;

RN [1]

RP SOURCE FROM N.A.

RA Knowles L., Otle D.;

RT "Phylogenetic analysis of montane grasshoppers from western North America (genus Melanoplus, Acrididae; Melanoplineae).";

RL Ann. Entomol. Soc. Amer. 93:421-431(2000).

CC -I- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3 AND COPPER B (BY SIMILARITY).

CC -I- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME C + 2 H(2)O.

CC -I- COFACTOR: HEMES A, A3, AND COPPER B (BY SIMILARITY).

CC -I- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.

CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL INNER MEMBRANE (BY SIMILARITY).

CC -I- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.

DR EMBL: AF229005; AAC35281.1; -.

DR InterPro: IPR000883; COX1.

DR Pfam: PF00115; COX1.2.

DR PRINTS: PRO1165; CYCOXIDASE1.

KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;

KW Respiratory chain; Transmembrane.

FT NON_TER 1 1

FT NON_TER 409 409

SO SEQUENCE 409 AA; 45084 MW; 0D177A05956CC655 CRC64;

Alignment Scores:

Pred. No.: 5.69e-07 Length: 409

Score: 122.00 Matches: 12

Percent Similarity: 54.22% Conservative: 33

Best Local Similarity: 14.46% Mismatches: 23

Query Match: 22.89% Indels: 15

Gaps: 3

Db: 8

US-09-880-711-328 (1-350) x 09G5C9 (1-409)

OY 124 TGCATGCCACAGCGCNGTATACATGCCTAGCAGNNNCAGCNCCTGACNNNTGCANACTGNC 183

Db 165 CysgIngluSerGlyIysIle-----GluSerPheGlyThrLeuGlyMetIle 180

OY 184 TNCACACATGATGATGATGCGTCGCGAATGCGCTGNCAGCATCAACTACTGCAG 243

Db 181 TyrAlaMetLeuSerIleGlyLeuMetGlyPheIleValITrPalAlaHisIleMetPhe--- 199

OY 244 ACCAANCAGGAGCTTNGAGACAGA-----ATGTAAT 276

Db 200 --ThrvAlcIyMeLAsValAsPthrArGAla*****AlaThrIleIleIle 218

OY 277 TANNACGACACTGTATNCTTTGCAANAAGAACTNTTCTTGANAATNTTTGACA 336

Db 219 AlaValProThrGlyIleIysValPheSerTrpLeuAlaThrLeuTyroGlyThrLysPhe 238

OY 337 GATTTCAGAT 345

Db 239 LysPheAsn 241

RESULT 14

08WJF5 PRELIMINARY: PRT: 460 AA.

AC 08WJF5:

DT 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE ATP synthase beta subunit (Fragment).

GN ATPB.

OS Thissmia rodwayi.

OC Chloroplast.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Dioscoreales; Thissmiaceae;

OX Thissmia.

OX NCBI_TaxID=147078;

RN [1]

RP SOURCE FROM N.A.

RA Caddick L.R., Kudall P.J., Wilkin P., Hedderston T.A.J., Chase M.W.;

RT "Phylogeny and circumscription of Dioscoreales based on combined analyses of morphological and molecular data.";

RT Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF308034; AAL37109.1; -.

DR InterPro: IPR000793; ATPase_a/b/c.

DR InterPro: IPR000154; ATPase_a/b/cenile.

DR InterPro: IPR004100; ATPase_a/bN.

DR Pfam: PF00006; ATP-synt_ab_1.

DR Pfam: PF0306; ATP-synt_ab_C_1.

DR Pfam: PF02874; ATP-synt_ab_N_1.

KW Chloroplast.

FT NON_TER 1 1

FT NON_TER 460 460

SO SEQUENCE 460 AA; 49784 MW; 025797491C7E9ECE CRC64;

Alignment Scores:

Pred. No.: 5.68e-07 Length: 460

Score: 122.00 Matches: 12

Percent Similarity: 61.54% Conservative: 36

Best Local Similarity: 15.38% Mismatches: 24

Query Match: 22.89% Indels: 6

Gaps: 2

Db: 8

US-09-880-711-328 (1-350) x 08WJF5 (1-460)

OY 133 AGAGGCTGTATATACATGCCTAGCAGNNNCAGCNCCTGACNNNTGCANACTGNCAGACA 192

Db 62 ArgGlyMetGluValIleAspThrGlyAlaProLeuSerValProValGlyIylarThr 81

OY 193 -----TGNATGATGATGCGTCGCGAATGCGCTGNCAGCATCAACTACTGCAGACC 246

Db 82 LeuGlyArgIlePheAsnValIleGlyIleProValAspAsnLeuGlyProValAspThr 101

OY 247 ANCAAGGAGCTTNGAGACAGA-----ATGTATTANNAACNACTACTCTTA 294

Db 102 AgyThrThrSerProIleIleIsarGserAlaProAlaPheIleGlnLeuAspThrLysLeu 121

OY 295 TNCCTTTGCAANAAGAACTNTTCTTGANAATNTTTGACAGATTTGCAATTA 348

Db 122 SerIlePheGluThrGlyIleIysValValAspLeuValaProTyraArgArg 139

RESULT 15

019816 PRELIMINARY: PRT: 505 AA.

AC 019816:

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Probable Intron maturase (Maturase K).

GN Allamanda cathartica (Yellow allamanda).

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OH nucleic - protein search, using frame_plus_n2p model

Run on: April 28, 2003, 15:35:24 : Search time 14 Seconds
(without alignments)
1471.146 Million cell updates/sec

Title: US-09-880-711-328

Perfect score: 533

Sequence: 1 atgagctcacagtactctca.....tttgcagatcttcaataaa 350

Scoring table:

BLOSUM62DX
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 5.0 , Fgapext 7.0
Delop 5.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p_model -DEV=xlp
-O=/cgn2_1/USPTC/spool/us09880711/runat_28042003_112140_888/app_query.fasta.1.519
-DB=Issued_Patents_AA -QFMT=fastcan -SUFFIX=n2p.fai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62dx -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=us09880711.ecgn.1.1.28.ernun.28042003_112140_888 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WATT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: Issued_Patents_AA:*
2: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backliles1.pep:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	131	25.9	400	4	US-09-076-137-42
C 2	129	24.2	731	1	US-08-158-232-35
C 3	129	25.5	731	1	US-08-158-232-35
C 4	125	24.2	731	1	US-08-304-626-35
C 5	129	25.5	731	1	US-08-304-626-35
C 6	129	24.2	731	2	US-08-611-928-35
C 7	129	25.5	731	2	US-08-611-928-35
C 8	129	24.2	731	3	US-09-173-891-35
C 9	129	25.5	731	3	US-09-173-891-35
C 10	128	25.3	411	2	US-08-568-459A-15
C 11	128	25.3	411	2	US-08-487-826B-31
C 12	128	25.3	411	4	US-09-210-288-19

13	128	24.0	715	4	US-09-076-137-41	Sequence 41, Appl
14	127	23.8	716	1	US-08-316-301A-41	Sequence 41, Appl
C 15	127	25.1	716	1	US-08-316-301A-41	Sequence 41, Appl
C 16	127	25.1	1040	4	US-08-974-549A-727	Sequence 727, App
C 17	126	24.9	401	1	US-08-316-301A-42	Sequence 42, Appl
C 18	126	24.9	716	4	US-09-076-137-41	Sequence 41, Appl
C 19	125	24.5	401	1	US-08-316-301A-42	Sequence 42, Appl
C 20	124	24.5	181	2	US-08-558-823-3	Sequence 3, Appl
C 21	124	24.3	400	4	US-09-076-137-42	Sequence 42, Appl
C 22	123.5	24.4	137	1	US-08-291-060B-2	Sequence 2, Appl
C 23	123	24.3	409	4	US-09-207-359B-46	Sequence 46, Appl
C 24	122	24.1	116	4	US-05-457-056-3	Sequence 3, Appl
C 25	122	22.9	1040	4	US-08-974-549A-727	Sequence 727, App
C 26	121	24.7	181	2	US-08-558-823-3	Sequence 3, Appl
C 27	121	22.7	405	4	US-09-207-359B-46	Sequence 46, Appl
C 28	121	22.7	461	1	US-08-196-989B-13	Sequence 13, Appl
C 29	121	22.7	451	2	US-08-760-936-13	Sequence 13, Appl
C 30	120	22.5	137	1	US-08-291-060B-2	Sequence 2, Appl
C 31	120	23.7	362	2	US-08-568-459A-18	Sequence 18, Appl
C 32	120	23.7	362	2	US-08-487-826B-30	Sequence 30, Appl
C 33	120	23.7	362	4	US-09-210-288-18	Sequence 18, Appl
C 34	120	23.7	411	2	US-08-568-459A-20	Sequence 20, Appl
C 35	120	23.7	411	2	US-08-487-826B-32	Sequence 32, Appl
C 36	120	23.7	411	4	US-09-210-288-20	Sequence 20, Appl
C 37	119	22.3	122	1	US-08-291-060B-3	Sequence 3, Appl
C 38	119	22.3	421	1	US-08-196-989B-12	Sequence 12, Appl
C 39	119	22.3	421	2	US-08-760-936-12	Sequence 12, Appl
C 40	119	22.5	820	4	US-09-391-104-7	Sequence 12, Appl
C 41	118	22.1	450	1	US-08-196-989B-11	Sequence 11, Appl
C 42	118	22.1	450	2	US-08-760-936-11	Sequence 11, Appl
C 43	118	23.3	461	1	US-08-196-989B-13	Sequence 13, Appl
C 44	118	23.3	461	2	US-08-760-936-13	Sequence 13, Appl
C 45	118	22.1	820	4	US-09-391-104-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1
US-09-076-137-42
Sequence 42, Application US/09076137B
Patent No. 6166195
GENERAL INFORMATION:
APPLICANT: Schuopt, Harry E.
APPLICANT: Schuopt, George E.
APPLICANT: Payne, Jewel M.
APPLICANT: Narva, Kenneth E.
TITLE OF INVENTION: No. 6166195el Nematode-Active Toxins and Genes which Code
FILE REFERENCE: MA-20CC02
CURRENT APPLICATION NUMBER: US/09/076,137B
CURRENT FILING DATE: 1998-05-12
EARLIER APPLICATION NUMBER: 08/316,301
EARLIER FILING DATE: 1994-09-30
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 42
LENGTH: 400
TYPE: PRT
ORGANISM: Generic Formula 11
US-09-076-137-42
Alignment Scores:
Pred. No.: 1176-08
Score: 131.00
Percent Similarity: 100.00%
Best Local Similarity: 83.33%
Query Match: 25.89%
DB: 4
Caps: 0
US-09-880-711-328 (1-350) x US-09-076-137-42 (1-400)
QY 165 GNGTCNNNCCTAGGCAT 148

Db 7 *****Prolyshis 12

RESULT 2

US-08-158-232-39
Sequence 39, Application US/08158232
Patent No. 5596071

GENERAL INFORMATION:

APPLICANT: Payne, Jewel
APPLICANT: Kennedy, M. Keith
APPLICANT: Randall, John Brooks
APPLICANT: Meier, Henry
APPLICANT: Uick, Heidi Jane
APPLICANT: Foncerrada, Luis
APPLICANT: Schnepf, H. Ernest
APPLICANT: Schwab, George E.
APPLICANT: Fu, Jenny

TITLE OF INVENTION: No. 5596071el Bacillus thuringiensis Toxins Active
TITLE OF INVENTION: Against Hymenopteran Pests
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:

ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA

ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/158,232
FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/887,980
FILING DATE: 22-MAY-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/797,645
FILING DATE: 25-NOV-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/703,977
FILING DATE: 22-MAY-1991

ATTORNEY/AGENT INFORMATION:

NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794

REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION: M/SCJ104.C1

TELEPHONE:

TELEPHONE: 904-375-8100

TELEFAX:

TELEFAX: 904-372-5800

INFORMATION FOR SEQ ID NO:

SEQUENCE CHARACTERISTICS:
LENGTH: 731 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: YES

US-08-158-232-39

Alignment Scores:

Pred. No.: 2.15e-08 Length: 731
Score: 129.00 Matches: 2
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 24.20% Indels: 0
DB: 1 Gaps: 0

US-09-880-711-328 (1-350) x US-08-158-232-39 (1-731)

OY 24 TACCC 29

Db 8 TyIPto 9

RESULT 3

US-08-158-232-39
Sequence 39, Application US/08158232
Patent No. 5596071

GENERAL INFORMATION:

APPLICANT: Payne, Jewel
APPLICANT: Kennedy, M. Keith
APPLICANT: Randall, John Brooks
APPLICANT: Meier, Henry
APPLICANT: Uick, Heidi Jane
APPLICANT: Foncerrada, Luis
APPLICANT: Schnepf, H. Ernest
APPLICANT: Schwab, George E.
APPLICANT: Fu, Jenny

TITLE OF INVENTION: No. 5596071el Bacillus thuringiensis Toxins Active
TITLE OF INVENTION: Against Hymenopteran Pests
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:

ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA

ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/158,232
FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/887,980
FILING DATE: 22-MAY-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/797,645
FILING DATE: 25-NOV-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/703,977
FILING DATE: 22-MAY-1991

ATTORNEY/AGENT INFORMATION:

NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794

REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION: M/SCJ104.C1

TELEPHONE:

TELEPHONE: 904-375-8100

TELEFAX:

TELEFAX: 904-372-5800

INFORMATION FOR SEQ ID NO:

SEQUENCE CHARACTERISTICS:
LENGTH: 731 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: YES

US-08-158-232-39

Alignment Scores:

Pred. No.: 2.15e-08 Length: 731
Score: 129.00 Matches: 3
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 75.00% Mismatches: 0
Query Match: 25.49% Indels: 0
DB: 1 Gaps: 0

US-09-880-711-328 (1-350) x US-08-158-232-39 (1-731)

OY 217 NCATCCGAGNG 206

```
||||:|||||
Db      7 ***TyrPro*** 10

RESULT 4
US-08-304-626-39
: Sequence 39, Application US/08304626
: Patent No. 5616495
: GENERAL INFORMATION:
: APPLICANT: Payne, Jewel M.
: APPLICANT: Kennedy, M. Keith
: APPLICANT: Randall, John Brooks
: APPLICANT: Meier, Henry
: APPLICANT: Uick, Heidi Jane
: APPLICANT: Poncetrada, Luis
: APPLICANT: Schepf, Harry E.
: APPLICANT: Schwab, George E.
: TITLE OF INVENTION: No. 5616495el Bacillus thuringiensis Isolates
: TITLE OF INVENTION: Active Against Hymenopteran Pests and Genes Encoding
: TITLE OF INVENTION: Hymenopteran-Active Toxins
: NUMBER OF SEQUENCES: 39
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: David R. Salimanchik
: STREET: 2121 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32609
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/304,626
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/887,980
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Salimanchik, David R.
: REGISTRATION NUMBER: 31,794
: REFERENCE/DOCKET NUMBER: M/SCJ 104
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 904-375-8100
: TELEFAX: 904-372-5800
: INFORMATION FOR SEQ ID NO: 39:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 741 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: HYPOTHEICAL: YES
: US-08-304-626-39

Alignment Scores:
Pred. No.: 2,15c-08 Length: 731
Score: 129.00 Matches: 2
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 24.20% Indels: 0
Gaps: 0
DB: 1

US-09-880-711-328 (1-350) x US-08-304-626-39 (1-731)

QY      24 TACCC 29
|||||
Db      8 TyrPro 5

RESULT 5
US-08-304-626-39
: Sequence 39, Application US/08304626
```

```
: Patent No. 5616495
: GENERAL INFORMATION:
: APPLICANT: Payne, Jewel M.
: APPLICANT: Kennedy, M. Keith
: APPLICANT: Randall, John Brooks
: APPLICANT: Meier, Henry
: APPLICANT: Uick, Heidi Jane
: APPLICANT: Poncetrada, Luis
: APPLICANT: Schepf, Harry E.
: APPLICANT: Schwab, George E.
: TITLE OF INVENTION: No. 5616495el Bacillus thuringiensis Isolates
: TITLE OF INVENTION: Active Against Hymenopteran Pests and Genes Encoding
: TITLE OF INVENTION: Hymenopteran-Active Toxins
: NUMBER OF SEQUENCES: 39
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: David R. Salimanchik
: STREET: 2121 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/304,626
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/887,980
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Salimanchik, David R.
: REGISTRATION NUMBER: 31,794
: REFERENCE/DOCKET NUMBER: M/SCJ 104
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 904-375-8100
: TELEFAX: 904-372-5800
: INFORMATION FOR SEQ ID NO: 39:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 741 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: HYPOTHEICAL: YES
: US-08-304-626-39

Alignment Scores:
Pred. No.: 2,15c-08 Length: 731
Score: 129.00 Matches: 3
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 75.00% Mismatches: 0
Query Match: 25.49% Indels: 0
Gaps: 0
DB: 0

US-09-880-711-328 (1-350) x US-08-304-626-39 (1-731)

QY      217 NCAATCTAGAG 205
||||:|||||
Db      7 ***TyrPro*** 10

RESULT 6
US-08-611-928-39
: Sequence 39, Application US/08611928
: Patent No. 5624792
: GENERAL INFORMATION:
: APPLICANT: Payne, Jewel
: APPLICANT: Kennedy, M. Keith
: APPLICANT: Randall, John Brooks
: APPLICANT: Meier, Henry
```

APPLICANT: Uick, Heidi Jane
APPLICANT: Foncerra, Luis
APPLICANT: Schepf, H. Ernest
APPLICANT: Schwab, George E.
APPLICANT: Fu, Jenny
TITLE OF INVENTION: No. 5824792el Bacillus thuringiensis Toxins Active
TITLE OF INVENTION: Against Hymenopteran Pests
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/611,928
FILING DATE: 06-MAR-1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/158,232
FILING DATE: 24-NOV-1993
APPLICATION NUMBER: US 07/887,980
FILING DATE: 22-MAY-1992
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/797,645
FILING DATE: 25-NOV-1991
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/703,977
FILING DATE: 22-MAY-1991
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: M/SCJ104.C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 731 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEICAL: YES
US-08-611-928-39

Alignment Scores:
Pred. No.: 2.15e-08 Length: 731
Score: 129.00 Matches: 2
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 24.20% Indels: 0
DB: 2 Gaps: 0
US-09-880-711-328 (1-350) x US-08-611-928-39 (1-731)

QY 24 TACCC 29
DB 8 TyrPro 9

RESULT 7
US-08-611-928-39
Sequence 39, Application US/08611928
Patent No. 5824792

GENERAL INFORMATION:
APPLICANT: Payne, Jewel
APPLICANT: Kennedy, M. Keith
APPLICANT: Randall, John Brooks
APPLICANT: Meier, Henry
APPLICANT: Uick, Heidi Jane
APPLICANT: Foncerra, Luis
APPLICANT: Schepf, H. Ernest
APPLICANT: Schwab, George E.
APPLICANT: Fu, Jenny
TITLE OF INVENTION: No. 5824792el Bacillus thuringiensis Toxins Active
TITLE OF INVENTION: Against Hymenopteran Pests
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/611,928
FILING DATE: 06-MAR-1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/158,232
FILING DATE: 24-NOV-1993
APPLICATION NUMBER: US 07/887,980
FILING DATE: 22-MAY-1992
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/797,645
FILING DATE: 25-NOV-1991
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/703,977
FILING DATE: 22-MAY-1991
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: M/SCJ104.C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 731 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEICAL: YES
US-08-611-928-39

Alignment Scores:
Pred. No.: 2.15e-08 Length: 731
Score: 129.00 Matches: 3
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 75.00% Mismatches: 0
Query Match: 25.49% Indels: 0
DB: 2 Gaps: 0
US-09-880-711-328 (1-350) x US-08-611-928-39 (1-731)

QY 217 NCATCCGAG 206
DB 7 ***TyrPro*** 10

RESULT 8
US-09-173-891-39
: Sequence 39, Application US/05173891
: Patent No. 6077937
: GENERAL INFORMATION:
: APPLICANT: Payne, Jewel
: APPLICANT: Kennedy, M. Keith
: APPLICANT: Randall, John Brooks
: APPLICANT: Meier, Henry
: APPLICANT: Uick, Heidi Jane
: APPLICANT: Fongertada, Luis
: APPLICANT: Schaepl, H. Ernest
: APPLICANT: Schwab, George E.
: APPLICANT: Fu, Jenny
: TITLE OF INVENTION: No. 6077937el Bacillus thuringiensis Toxins Active
: TITLE OF INVENTION: Against Hymenopteran Pests
: NUMBER OF SEQUENCES: 51
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: David R. Saliwanchik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/173,891
: FILING DATE:
: CLASSIFICATION:
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: 08/158,232
: FILING DATE:
: APPLICATION NUMBER: US 07/887,980
: FILING DATE: 22-MAY-1992
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 07/797,645
: FILING DATE: 25-NOV-1991
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 07/703,977
: FILING DATE: 22-MAY-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Saliwanchik, David R.
: REGISTRATION NUMBER: 31,794
: REFERENCE/DOCKET NUMBER: M/SCJ104.C1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 904-375-8100
: TELEFAX: 904-372-5800
: INFORMATION FOR SEQ ID NO: 39:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 731 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: HYPOTHEICAL: YES
: US-09-173-891-39
Alignment Scores:
Pred. No.: 2,15e-08 Length: 731
Score: 129.00 Matches: 2
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 24.20% Indels: 0
DB: 3 Gaps: 0
US-09-880-711-328 (1-350) x US-09-173-891-39 (1-731)
QY 24 TACCC 29

Db 8 11111
8 tyrPro 9
RESULT 9
US-09-173-891-39
: Sequence 39, Application US/09173891
: Patent No. 6077937
: GENERAL INFORMATION:
: APPLICANT: Payne, Jewel
: APPLICANT: Kennedy, M. Keith
: APPLICANT: Randall, John Brooks
: APPLICANT: Meier, Henry
: APPLICANT: Uick, Heidi Jane
: APPLICANT: Fongertada, Luis
: APPLICANT: Schaepl, H. Ernest
: APPLICANT: Schwab, George E.
: APPLICANT: Fu, Jenny
: TITLE OF INVENTION: No. 6077937el Bacillus thuringiensis Toxins Active
: TITLE OF INVENTION: Against Hymenopteran Pests
: NUMBER OF SEQUENCES: 51
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: David R. Saliwanchik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/173,891
: FILING DATE:
: CLASSIFICATION:
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: 08/158,232
: FILING DATE:
: APPLICATION NUMBER: US 07/887,980
: FILING DATE: 22-MAY-1992
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 07/797,645
: FILING DATE: 25-NOV-1991
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 07/703,977
: FILING DATE: 22-MAY-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Saliwanchik, David R.
: REGISTRATION NUMBER: 31,794
: REFERENCE/DOCKET NUMBER: M/SCJ104.C1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 904-375-8100
: TELEFAX: 904-372-5800
: INFORMATION FOR SEQ ID NO: 39:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 731 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: HYPOTHEICAL: YES
: US-09-173-891-39
Alignment Scores:
Pred. No.: 2,15e-08 Length: 731
Score: 129.00 Matches: 3
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 75.00% Mismatches: 0
Query Match: 25.49% Indels: 0
DB: 3 Gaps: 0
US-09-880-711-328 (1-350) x US-09-173-891-39 (1-731)

OY 217 NCATCCGANG 206
Db 7 ***TyPro*** 10

RESULT 10
US-08-568-459A-19
Sequence 19, Application US/08568459A

Patent No. 5845306

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.

APPLICANT: Chitnis, Chetan

APPLICANT: Miller, Louis H.

APPLICANT: Peterson, David S.

APPLICANT: Su, Xin-zhaun

APPLICANT: Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe Martens Olson & Bear

STREET: 620 Newport Center Drive 16th Floor

CITY: Newport Beach

STATE: California

COUNTRY: US

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/568,459A

FILING DATE: 07-DEC-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Israelson, Ned

REGISTRATION NUMBER: 29,655

REFERENCE/DOCKET NUMBER: NIH21.001CP1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 235-8550

TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 411 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: Internal

ORIGINAL SOURCE:

US-08-568-459A-19

Alignment Scores:

Pred. No.: 2.8e-08 Length: 411

Score: 128.00 Matches: 5

Percent Similarity: 88.89% Conservative: 3

Best Local Similarity: 55.56% Mismatches: 1

Query Match: 25.30% Indels: 0

DB: 2 Gaps: 0

US-09-880-711-328 (1-350) x US-08-568-459A-19 (1-411)

OY 148 TGTATACAGNGCCTCTGGCATGCAGT 122

Db 278 CysIleSerCysLeuTyAlaCysasn 286

RESULT 11

US-08-487-826B-31

Sequence 31, Application US/08487826B

Patent No. 5993827

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.

APPLICANT: Chitnis, Chetan

APPLICANT: Miller, Louis H.

APPLICANT: Peterson, David S.

APPLICANT: Su, Xin-zhaun

APPLICANT: Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 45

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe Martens Olson & Bear

STREET: 620 Newport Center Drive 16th Floor

CITY: Newport Beach

STATE: California

COUNTRY: US

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/487,826B

FILING DATE: 10-SEP-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Israelson, Ned

REGISTRATION NUMBER: 29,655

REFERENCE/DOCKET NUMBER: NIH21.001CP1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 235-8550

TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 31:

SEQUENCE CHARACTERISTICS:

LENGTH: 411 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: Internal

ORIGINAL SOURCE:

US-08-487-826B-31

Alignment Scores:

Pred. No.: 2.8e-08 Length: 411

Score: 128.00 Matches: 5

Percent Similarity: 88.89% Conservative: 3

Best Local Similarity: 55.56% Mismatches: 1

Query Match: 25.30% Indels: 0

DB: 2 Gaps: 0

US-09-880-711-328 (1-350) x US-08-487-826B-31 (1-411)

OY 148 TGTATACAGNGCCTCTGGCATGCAGT 122

Db 278 CysIleSerCysLeuTyAlaCysasn 286

RESULT 12

US-09-210-288-19

Sequence 19, Application US/09210288

Patent No. 6392026

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.

APPLICANT: Chitnis, Chetan

APPLICANT: Miller, Louis H.

APPLICANT: Peterson, David S.

APPLICANT: Su, Xin-zhaun

APPLICANT: Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbie Martens Olson & Neat
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1PMDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 411 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
US-09-210-288-19

Alignment Scores:
Pred. No.: 2.8e-08 Length: 411
Score: 128.00 Matches: 5
Percent Similarity: 88.89% Conservative: 3
Best Local Similarity: 55.56% Mismatches: 1
Query Match: 25.30% Indels: 0
DB: 4 Gaps: 0

US-09-880-711-323 (1-350) x US-09-210-288-19 (1-411)

QY 148 TGTATACAGNCCCTGTGGCATGCACT 122
||||| :||||:|||||:
DB 278 CysIeserCysIenTyraIacysasn 286

RESULT 13
US-09-076-137-41

Sequence 41, Application US/09076137B

Patent No. 6165195

GENERAL INFORMATION:
APPLICANT: Schmepl, Harry E.
APPLICANT: Scwab, George E.
APPLICANT: Payne, Jewel M.
APPLICANT: Narva, Kenneth E.
APPLICANT: Poncetrada, Luis
TITLE OF INVENTION: No. 6165195e1 Nematode-Active Toxins and Genes Which Code
FILE REFERENCE: MA-20CCCD2
CURRENT APPLICATION NUMBER: US/09/076,137B
CURRENT FILING DATE: 1994-05-12
EARLIER APPLICATION NUMBER: 08/316,301
EARLIER FILING DATE: 1994-09-30
NUMBER OF SEQ ID NOS: 42
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 41
LENGTH: 716
TYPE: PRT

ORGANISM: Generic Formula 1
US-09-076-137-41

Alignment Scores:
Pred. No.: 2.86e-08 Length: 716
Score: 128.00 Matches: 2
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 56.67% Mismatches: 0
Query Match: 24.02% Indels: 0
DB: 4 Gaps: 0

US-09-880-711-328 (1-350) x US-09-076-137-41 (1-716)

QY 98 TGGCCGCTT 106
|||||:
DB 622 CysArgTyr 624

RESULT 14
US-08-316-301A-41

Sequence 41, Application US/08316301A

Patent No. 5753492

GENERAL INFORMATION:
APPLICANT: Schmepl, Harry E.
APPLICANT: Schwab, George E.
APPLICANT: Payne, Jewel M.
APPLICANT: Narva, Kenneth E.
APPLICANT: Poncetrada, Luis
TITLE OF INVENTION: No. 5753492e1 Nematode-Active Toxins and Genes
TITLE OF INVENTION: Which Code Therefor
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/316,301A
FILING DATE: 30-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/871,510
FILING DATE: 23-APR-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/693,018
FILING DATE: 03-MAY-1991
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/565,544
FILING DATE: 10-AUG-1990
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/084,653
FILING DATE: 12-AUG-1987
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/830,050
FILING DATE: 31-JAN-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lloyd, Jeff
REGISTRATION NUMBER: 35,589
REFERENCE/DOCKET NUMBER: MA20CCCD1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5890

```

: INFORMATION FOR SEQ ID NO: 41:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 716 amino acids
:   TYPE: amino acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-316-301A-41

Alignment Scores:
Pred. No.:      3,82e-08      Length: 716
Score:          127.00      Matches: 2
Percent Similarity: 100.00%      Conservative: 1
Best Local Similarity: 66.67%      Mismatches: 0
Query Match:      23.83%      Indels: 0
DB:               1      Gaps: 0

US-09-880-711-328 (1-350) x US-08-316-301A-41 (1-716).
QY  98  TGCCGCTT 106
    |||||:::
Db  622  CysArgTyr 624

RESULT 15
US-08-316-301A-41
: Sequence 41, Application.US/08316301A
: Patent No. 5753492
: GENERAL INFORMATION:
:   APPLICANT: Schmepl, Harry E.
:   APPLICANT: Schwab, George E.
:   APPLICANT: Payne, Jewel M.
:   APPLICANT: Narva, Kenneth E.
:   APPLICANT: Fonzerrada, Luis
:   TITLE OF INVENTION: No. 5753492el Nematode-Active Toxins and Genes
:   TITLE OF INVENTION: Which Code Therefor
:   NUMBER OF SEQUENCES: 42
:   CORRESPONDENCE ADDRESS:
:     ADDRESSEE: Saliwanchik & Saliwanchik
:     STREET: 2421 N.W. 41st Street, Suite A-1
:     CITY: Gainesville
:     STATE: FL
:     COUNTRY: USA
:     ZIP: 32606
:   COMPUTER READABLE FORM:
:     MEDIUM TYPE: Floppy disk
:     COMPUTER: IBM PC compatible
:     OPERATING SYSTEM: PC-DOS/MS-DOS
:     SOFTWARE: Patent Release #1.0, Version #1.25
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/08/316,301A
:     FILING DATE: 30-SEP-1994
:     CLASSIFICATION: 435
:     PRIOR APPLICATION DATA:
:       APPLICATION NUMBER: 07/871,510
:       FILING DATE: 23-APR-1992
:       CLASSIFICATION: 435
:     PRIOR APPLICATION DATA:
:       APPLICATION NUMBER: 07/693,018
:       FILING DATE: 03-MAY-1991
:       CLASSIFICATION: 435
:     PRIOR APPLICATION DATA:
:       APPLICATION NUMBER: 07/565,544
:       FILING DATE: 10-AUG-1990
:       CLASSIFICATION: 435
:     PRIOR APPLICATION DATA:
:       APPLICATION NUMBER: 07/084,653
:       FILING DATE: 12-AUG-1987
:       CLASSIFICATION: 435
:     PRIOR APPLICATION DATA:
:       APPLICATION NUMBER: 07/830,050
:       FILING DATE: 31-JAN-1992
:       CLASSIFICATION: 435
:     ATTORNEY/AGENT INFORMATION:

```

```

NAME: Lloyd, Jeff
REGISTRATION NUMBER: 35,589
REFERENCE/DOCKET NUMBER: MA20CCCD1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 41:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 716 amino acids
:   TYPE: amino acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-316-301A-41

Alignment Scores:
Pred. No.:      3,82e-08      Length: 716
Score:          127.00      Matches: 3
Percent Similarity: 100.00%      Conservative: 1
Best Local Similarity: 75.00%      Mismatches: 0
Query Match:      25.10%      Indels: 0
DB:               1      Gaps: 0

US-09-880-711-328 (1-350) x US-08-316-301A-41 (1-716)
QY  23  AGGTGAGGTAC 12
    ::|||
Db  621  ArgCysArgTyr 624

```

Search completed: April 28, 2003, 15:39:12
Job time : 17 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OH nucleic - protein search, using frame_plus_n2p model

Run on: April 28, 2003, 15:39:10 ; Search time 19 Seconds

(without alignments)
2552.151 Million cell updates/sec

Title: US-09-880-711-328

Perfect score: 533
Sequence: 1 atgagctcagctacgtcga.....ltgagagatllcaataaaa 350

Scoring table: BLOSUM62DX

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 301932 seqs, 80125803 residues

Total number of hits satisfying chosen parameters: 603864

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0k
Maximum Match 100k
Listing first 45 summaries

Command line parameters:
-MODEL=frame+2p.model -DPV=xlp
-Q=/cgn2_1/usprto.spool/US09880711/runat.28042003_112140_904/app_query.fasta.1.519
-DB=published_Applications_AA -QPMT=fastan -SUFEX=n2p.rapb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blotsum62dx
-TRANS=human40.csl -LIST=45 -DOCALLIGN=200 -THR_SCORE=E-PCT -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09880711_6CGN_1_1_17_6runat.28042003_112140_904
-KCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NRC_SCORES=0 -NAIT -LONGLOC
-DEV_TTMOUT=120 -WARN_TTMOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6
-FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published_Applications_AA:*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_MIN_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCUUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	132.5	24.9	163	10	US-09-867-550-834
2	128	24.0	232	10	US-09-854-122-1
3	128	25.3	411	9	US-10-153-273-19
4	126	23.6	1208	9	US-10-101-812-11

C 5	126	24.9	1208	9	US-10-101-812-11	Sequence 11, Appl
C 6	125	24.7	513	9	US-09-051-013-2	Sequence 2, Appl1
C 7	125	24.7	1206	9	US-10-101-812-9	Sequence 2, Appl1
C 8	125	24.7	1206	9	US-10-101-662A-14	Sequence 14, Appl
C 9	124	23.3	175	10	US-09-796-338A-26	Sequence 26, Appl
C 10	124	23.3	413	9	US-10-154-155-1	Sequence 1, Appl1
C 11	124	23.3	513	9	US-09-051-013-2	Sequence 2, Appl1
C 12	124	24.5	1210	9	US-10-101-812-6	Sequence 2, Appl1
C 13	124	24.5	1210	9	US-10-101-652A-13	Sequence 13, Appl
C 14	123	24.3	175	10	US-09-908-805B-59	Sequence 59, Appl
C 15	123	24.3	175	10	US-09-796-338A-26	Sequence 26, Appl
C 16	123	24.1	261	10	US-09-516-790-37	Sequence 37, Appl
C 17	122	24.1	116	9	US-10-139-583-3	Sequence 3, Appl1
C 18	121	22.7	1206	9	US-10-101-812-9	Sequence 9, Appl1
C 19	121	22.7	1206	9	US-10-101-652A-14	Sequence 14, Appl
C 20	120	22.5	128	10	US-09-508-805B-59	Sequence 59, Appl
C 21	120	23.7	362	9	US-10-153-273-18	Sequence 18, Appl
C 22	120	23.7	364	10	US-09-051-755-18	Sequence 18, Appl
C 23	120	22.5	411	9	US-10-153-273-20	Sequence 20, Appl
C 24	120	22.5	1208	9	US-10-101-662A-12	Sequence 12, Appl
C 25	120	23.7	1208	9	US-10-101-652A-12	Sequence 12, Appl
C 26	120	22.5	1333	9	US-09-972-115A-2	Sequence 2, Appl1
C 27	119	23.5	261	10	US-09-916-790-37	Sequence 37, Appl
C 28	119	23.5	270	10	US-09-815-915-17	Sequence 17, Appl
C 29	119	23.5	270	10	US-09-815-915-17	Sequence 17, Appl
C 30	119	22.3	270	10	US-09-910-150-37	Sequence 37, Appl
C 31	119	23.5	270	10	US-09-910-150-37	Sequence 37, Appl
C 32	119	22.3	270	10	US-09-834-456A-6	Sequence 6, Appl1
C 33	119	23.5	270	10	US-09-834-456A-6	Sequence 6, Appl1
C 34	119	22.3	270	10	US-09-515-806-5	Sequence 5, Appl1
C 35	119	22.3	270	10	US-09-515-806-5	Sequence 5, Appl1
C 36	115	22.3	270	10	US-09-860-352A-4	Sequence 4, Appl1
C 37	119	22.3	1034	10	US-09-858-754-5	Sequence 5, Appl1
C 38	119	22.3	1267	9	US-09-972-115A-4	Sequence 4, Appl1
C 39	119	22.3	1338	9	US-10-029-413A-10	Sequence 10, Appl
C 40	118	22.1	1210	9	US-10-101-812-6	Sequence 6, Appl1
C 41	118	22.1	1210	9	US-10-101-662A-13	Sequence 13, Appl
C 42	117	23.1	232	10	US-09-854-122-1	Sequence 1, Appl1
C 43	117	23.1	270	10	US-09-860-352A-4	Sequence 4, Appl1
C 44	117	22.0	364	10	US-09-051-755-18	Sequence 18, Appl
C 45	117	23.1	413	9	US-10-154-155-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-09-867-550-834
Sequence 834, Application US/09867550
Patent No. US20020082206A1
GENERAL INFORMATION:
APPLICANT: Leach, Martin D.
APPLICANT: Mehrtz, Rued,
APPLICANT: Conley, Pamela
APPLICANT: Law, Debbie
TITLE OF INVENTION: Topnet, James
TITLE OF INVENTION: Thereby
FILE REFERENCE: 21402-013 (Cura-313)
CURRENT APPLICATION NUMBER: US/09/867,550
CURRENT FILING DATE: 2001-09-20
PRIOR APPLICATION NUMBER: USSN 60/208,427
PRIOR FILING DATE: 2000-05-30
NUMBER OF SEQ ID NOS: 2125
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 834
LENGTH: 163
TYPE: PRT
ORGANISM: Homo sapiens
US-09-867-550-834
Alignment Scores: 1.33e-07 Length: 163
Pred. No.: 132.50 Matches: 29

SEQUENCE 19, APPLICATION US/101332/3

TITLE OF INVENTION: Pharmaceuticals and Methods

```

: TITLE OF INVENTION: Screening Methods Thereof
: FILE REFERENCE: 20363-014
: CURRENT APPLICATION NUMBER: US/10/101,812
: CURRENT FILING DATE: 2002-06-07
: PRIOR APPLICATION NUMBER: 60/277,425
: PRIOR FILING DATE: 2001-03-20
: PRIOR APPLICATION NUMBER: 60/277,431
: PRIOR FILING DATE: 2001-03-20
: PRIOR APPLICATION NUMBER: 60/277,440
: PRIOR FILING DATE: 2001-03-20
: PRIOR APPLICATION NUMBER: 60/332,493
: PRIOR FILING DATE: 2001-11-09
: PRIOR APPLICATION NUMBER: 60/332,334
: PRIOR FILING DATE: 2001-11-09
: PRIOR APPLICATION NUMBER: 60/345,200
: PRIOR FILING DATE: 2001-11-09
: PRIOR APPLICATION NUMBER: 60/345,131
: PRIOR FILING DATE: 2001-12-20
: PRIOR APPLICATION NUMBER: 60/342,598
: PRIOR FILING DATE: 2001-12-20
: PRIOR APPLICATION NUMBER: 60/345,132
: PRIOR FILING DATE: 2001-12-20
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 11
: LENGTH: 1208
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Consensus
: OTHER INFORMATION: Binding Peptide
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (1)..(600)
: OTHER INFORMATION: A peptide of 1 to 600 amino acids long wherein Xaa
: OTHER INFORMATION: Is any amino acid.
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (601)
: OTHER INFORMATION: Wherein Xaa is Gly, Ala, Val, Leu, Ile, Pro, Met,
: OTHER INFORMATION: Phe, or Trp.
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (603)
: OTHER INFORMATION: Wherein Xaa is Gly, Ala, Val, Leu, Ile, Pro, Met,
: OTHER INFORMATION: Phe, or Trp.
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (605)
: OTHER INFORMATION: Wherein Xaa is Ser, Thr, or Tyr.
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (606)..(608)
: OTHER INFORMATION: Wherein Xaa is Gly, Ala, Val, Leu, Ile, Pro, Met,
: OTHER INFORMATION: Phe, or Trp.
: US-10-101-812-11

Alignment Scores:
Pred. No.: 7.27e-07 Length: 1208
Score: 126.00 Matches: 3
Percent Similarity: 100.00% Conservative: 0
Best local Similarity: 100.00% Mismatches: 0
Query Match: 23.64% Indels: 0
DB: 9 Gaps: 0

US-09-880-711-328 (1-350) x US-10-101-812-11 (1-1208):
OY 291 CTTATNCT 299
Db 602 Leu***Pro 604

RESULT 5
US-10-101-812-11
```

```

: Sequence 11, Application US/10101812
: Publication No. US20020192737A1
: GENERAL INFORMATION:
: APPLICANT: Kaelin Jr., William G
: TITLE OF INVENTION: Pharmaceuticals and Methods for Treating Hypoxia and
: FILE REFERENCE: 20363-014
: CURRENT APPLICATION NUMBER: US/10/101,812
: CURRENT FILING DATE: 2002-06-07
: PRIOR APPLICATION NUMBER: 60/277,425
: PRIOR FILING DATE: 2001-03-20
: PRIOR APPLICATION NUMBER: 60/277,431
: PRIOR FILING DATE: 2001-03-20
: PRIOR APPLICATION NUMBER: 60/277,440
: PRIOR FILING DATE: 2001-03-20
: PRIOR APPLICATION NUMBER: 60/332,493
: PRIOR FILING DATE: 2001-11-09
: PRIOR APPLICATION NUMBER: 60/332,334
: PRIOR FILING DATE: 2001-11-09
: PRIOR APPLICATION NUMBER: 60/345,200
: PRIOR FILING DATE: 2001-11-09
: PRIOR APPLICATION NUMBER: 60/345,131
: PRIOR FILING DATE: 2001-12-20
: PRIOR APPLICATION NUMBER: 60/342,598
: PRIOR FILING DATE: 2001-12-20
: PRIOR APPLICATION NUMBER: 60/345,132
: PRIOR FILING DATE: 2001-12-20
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 11
: LENGTH: 1208
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Consensus
: OTHER INFORMATION: Binding Peptide
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (1)..(600)
: OTHER INFORMATION: A peptide of 1 to 600 amino acids long wherein Xaa
: OTHER INFORMATION: Is any amino acid.
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (601)
: OTHER INFORMATION: Wherein Xaa is Gly, Ala, Val, Leu, Ile, Pro, Met,
: OTHER INFORMATION: Phe, or Trp.
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (603)
: OTHER INFORMATION: Wherein Xaa is Gly, Ala, Val, Leu, Ile, Pro, Met,
: OTHER INFORMATION: Phe, or Trp.
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (605)
: OTHER INFORMATION: Wherein Xaa is Ser, Thr, or Tyr.
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (606)..(608)
: OTHER INFORMATION: Wherein Xaa is Gly, Ala, Val, Leu, Ile, Pro, Met,
: OTHER INFORMATION: Phe, or Trp.
: US-10-101-812-11

Alignment Scores:
Pred. No.: 7.27e-07 Length: 1208
Score: 126.00 Matches: 4
Percent Similarity: 100.00% Conservative: 1
Best local Similarity: 80.00% Mismatches: 0
Query Match: 24.90% Indels: 0
DB: 9 Gaps: 0

US-09-880-711-328 (1-350) x US-10-101-812-11 (1-1208)
```

LENGTH: 1200

PRIOR FILING DATE

PRIOR FILING DATE

SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 14
LENGTH: 1206
TYPE: PRP
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)..(600)
OTHER INFORMATION: A peptide of 1 to 600 amino acids wherein Xaa is
OTHER INFORMATION: any amino acid.
FEATURE:
NAME/KEY: VARIANT
LOCATION: (602)
OTHER INFORMATION: wherein Xaa is any amino acid.
FEATURE:
NAME/KEY: VARIANT
LOCATION: (605)
OTHER INFORMATION: wherein Xaa is any amino acid.
FEATURE:
NAME/KEY: VARIANT
LOCATION: (607)..(1206)
OTHER INFORMATION: A peptide of 1 to 600 amino acids wherein Xaa is
OTHER INFORMATION: any amino acid.
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: consensus
US-10-101-662A-14

Alignment Scores:
Pred. No.: 9,54e-07 Length: 1206
Score: 125.00 Matches: 3
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 75.00% Mismatches: 0
Query Match: 24.70% Indels: 0
DB: 9 Gaps: 0

US-09-880-711-328 (1-350) x US-10-101-662A-14 (1-1206)

QY 335 CTCGACNATTNT 324

DB 603 LeuLys**Leu 606

RESULT 9
US-09-796-338A-26
Sequence 26, Application US/09796338A
PATENT NO. US:0020061522A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
TITLE OF INVENTION: 1983, 52881, 2398, 45449, 50289, AND
TITLE OF INVENTION: 52872, NOVEL G PROTEIN-COUPLED RECEPTORS AND USES THEREFOR
FILE REFERENCE: 10448-020001
CURRENT APPLICATION NUMBER: US/09/796,338A
CURRENT FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 60/186,059
PRIOR FILING DATE: 2000-02-29
NUMBER OF SEQ ID NOS: 26
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 26
LENGTH: 175
TYPE: PRP
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: exemplary motif
NAME/KEY: VARIANT
LOCATION: 1-4, 6-53, 55-66, 68-137, 139-144, 146, 147, 150-170, 172,
LOCATION: 173, 175
OTHER INFORMATION: Xaa = any amino acid
NAME/KEY: VARIANT
LOCATION: 149
OTHER INFORMATION: Xaa = any aromatic amino acid
US-09-796-338A-26

Alignment Scores:

Pred. No.: 1.33e-06 Length: 175
Score: 124.00 Matches: 2
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 23.25% Indels: 0
DB: 10 Gaps: 0

US-09-880-711-328 (1-350) x US-09-796-338A-26 (1-175)

QY 138 CAGCTGT 143

DB 4 ***Cys 5

RESULT 10
US-10-194-155-1
Sequence 1, Application US/10194155
Publication No. US20030045679A1
GENERAL INFORMATION:
APPLICANT: Crawford, Dana R
TITLE OF INVENTION: Calcineurin Modulators
FILE REFERENCE: 19705-005
CURRENT APPLICATION NUMBER: US/10/194,155
CURRENT FILING DATE: 2002-07-12
PRIOR APPLICATION NUMBER: 60/305,202
PRIOR FILING DATE: 2001-07-13
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 413
TYPE: PRP
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Calcineurin
OTHER INFORMATION: modulator
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)..(200)
OTHER INFORMATION: wherein Xaa is any amino acid, or no amino acid,
OTHER INFORMATION: such that Xaa 1-200 and Xaa 205-405 combine to
OTHER INFORMATION: represent 200 amino acids total.
FEATURE:
NAME/KEY: VARIANT
LOCATION: (214)..(413)
OTHER INFORMATION: wherein Xaa is any amino acid, or no amino acid,
OTHER INFORMATION: such that Xaa 1-200 and Xaa 205-405 combine to
OTHER INFORMATION: represent 200 amino acids total.
US-10-194-155-1

Alignment Scores:
Pred. No.: 1.3e-06 Length: 413
Score: 124.00 Matches: 4
Percent Similarity: 80.00% Conservative: 0
Best Local Similarity: 80.00% Mismatches: 1
Query Match: 23.26% Indels: 0
DB: 9 Gaps: 0

US-09-880-711-328 (1-350) x US-10-194-155-1 (1-413)

QY 81 CAGCTCTTCGCTGT 95

DB 208 ProAlaSerProPro 212

RESULT 11
US-09-051-013-2
Sequence 2, Application US/09051013
Publication No. US20020188103A1
GENERAL INFORMATION:
APPLICANT: Basior, Timothy H.
TITLE OF INVENTION: CHIMERIC DNA-BINDING/DNA METHYLTRANSFERASE NUCLEIC ACID
FILE REFERENCE: 48075/B/PCT/US
CURRENT APPLICATION NUMBER: US/09/051,013
CURRENT FILING DATE: 1998-10-09

NUMBER OF SEQ ID NOS: 8
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 2
LENGTH: 513
TYPE: PRT
ORGANISM: Spiroplasma
FEATURE:
OTHER INFORMATION: X in this sequence is unknown
US-09-051-013-2

Alignment Scores:
Pred. No.: 1.29e-06 Length: 513
Score: 124.00 Matches: 4
Percent Similarity: 76.47% Conservative: 9
Best Local Similarity: 23.53% Mismatches: 4
Query Match: 23.26% Indels: 0
DB: 9 Gaps: 0

US-09-880-711-328 (1-350) x US-09-051-013-2 (1-513)

Qy 247 ANCNAGGAGCTTNNAGACAGATGTATTANNACNACNACTTATNC 297
Db 1 MetArgGlySerHisHisHisHisHisGlyIleCysThrThrMetSer 17

RESULT 12
US-10-101-812-6
Sequence 6, Application US/10101812.
Publication No. US20020192737A1
GENERAL INFORMATION:
APPLICANT: Kaelin Jr., William G
TITLE OF INVENTION: Pharmaceuticals and Methods for Treating Hypoxia and
FILE REFERENCE: 20363-014
CURRENT APPLICATION NUMBER: US/10/101,812
CURRENT FILING DATE: 2002-06-07
PRIOR APPLICATION NUMBER: 60/277,425
PRIOR FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: 60/277,431
PRIOR FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: 60/277,440
PRIOR FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: 60/332,493
PRIOR FILING DATE: 2001-11-09
PRIOR APPLICATION NUMBER: 60/332,334
PRIOR FILING DATE: 2001-11-09
PRIOR APPLICATION NUMBER: 60/345,200
PRIOR FILING DATE: 2001-11-09
PRIOR APPLICATION NUMBER: 60/345,131
PRIOR FILING DATE: 2001-12-20
PRIOR APPLICATION NUMBER: 60/342,598
PRIOR FILING DATE: 2001-12-20
PRIOR APPLICATION NUMBER: 60/345,132
PRIOR FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 6
LENGTH: 1210
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Consensus
OTHER INFORMATION: Target Peptide
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)..(600)
OTHER INFORMATION: A peptide of 1 to 600 amino acids wherein Xaa is
OTHER INFORMATION: any amino acid.
FEATURE:
NAME/KEY: VARIANT
LOCATION: (611)..(1210)
OTHER INFORMATION: A peptide of 1 to 600 amino acids wherein Xaa is
OTHER INFORMATION: any amino acid.

US-10-101-812-6

Alignment Scores:
Pred. No.: 1.25e-06 Length: 1210
Score: 124.00 Matches: 4
Percent Similarity: 100.00% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 0
Query Match: 24.51% Indels: 0
DB: 9 Gaps: 0

US-09-880-711-328 (1-350) x US-10-101-812-6 (1-1210)

Qy 178 NTCCANNNGTCAGGNGTGNNNCC 155
Db 599 ****GlySerClyIlePheLeu 606

RESULT 13
US-10-101-662A-13
Sequence 13, Application US/10101662A
Publication No. US20030022198A1
GENERAL INFORMATION:
APPLICANT: Kaelin Jr., William G
APPLICANT: Livingston, David A
TITLE OF INVENTION: Light Emitting Fusion Proteins and Diagnostic and
FILE REFERENCE: 20363-009
CURRENT APPLICATION NUMBER: US/10/101,662A
CURRENT FILING DATE: 2002-03-19
PRIOR APPLICATION NUMBER: 60/277,425
PRIOR FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: 60/277,431
PRIOR FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: 60/277,440
PRIOR FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: 60/332,493
PRIOR FILING DATE: 2001-11-09
PRIOR APPLICATION NUMBER: 60/345,131
PRIOR FILING DATE: 2001-12-20
PRIOR APPLICATION NUMBER: 60/342,598
PRIOR FILING DATE: 2001-12-20
PRIOR APPLICATION NUMBER: 60/345,132
PRIOR FILING DATE: 2001-12-20
PRIOR APPLICATION NUMBER: 60/332,334
PRIOR FILING DATE: 2001-11-09
PRIOR APPLICATION NUMBER: 60/345,200
PRIOR FILING DATE: 2001-11-09
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 13
LENGTH: 1210
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Consensus
OTHER INFORMATION: Target Peptide
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)..(600)
OTHER INFORMATION: A peptide of 1 to 600 amino acids wherein Xaa is
OTHER INFORMATION: any amino acid.
FEATURE:
NAME/KEY: VARIANT
LOCATION: (611)..(1210)
OTHER INFORMATION: A peptide of 1 to 600 amino acids wherein Xaa is
OTHER INFORMATION: any amino acid.
US-10-101-662A-13

Alignment Scores:
Pred. No.: 1.25e-06 Length: 1210
Score: 124.00 Matches: 4
Percent Similarity: 100.00% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 0


```

Query Match:          24.51%      Indels:      0
DB:                   9           Gaps:         0

US-09-880-711-328 (1-350) x US-10-101-662A-13 (1-1210)

Qy      178 NTCANNNGTCAGCGNGTGNMNC 155
      |||||:|||||:|||||:
DB      599 *****GlySerGlyIlePheLeu 606

RESULT 14
US-09-908-805B-59
: Sequence 59, Application US/0990805B
: Patent No. US26020147307A1
: GENERAL INFORMATION:
: APPLICANT: HILTON, Douglas J
: APPLICANT: ALEXANDER, Warren S
: APPLICANT: VINAY, Elizabeth M
: APPLICANT: WILSON, Tracey A
: APPLICANT: RICHARDSON, Rachael T
: APPLICANT: STARR, Robyn
: APPLICANT: NICHOLSON, Sandra E
: APPLICANT: METCALF, Donald
: APPLICANT: NICOLA, Nicos A
: TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC AGENTS
: FILE REFERENCE: 109762
: CURRENT APPLICATION NUMBER: US/09/908.805B
: PRIORITY FILING DATE: 2001-07-19
: PRIOR APPLICATION NUMBER: 08/962,560
: NUMBER OF SEQ ID NOS: 81
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 59
: LENGTH: 128
: TYPE: PRT
: ORGANISM: Mus musculus
: FEATURE:
: NAME/KEY: UNSURE
: LOCATION: (1)
: OTHER INFORMATION: Xaa is Leu, Ile, Val, Met, Ala or Pro
: NAME/KEY: UNSURE
: LOCATION: (2)
: OTHER INFORMATION: Xaa is any amino acid residue
: NAME/KEY: UNSURE
: LOCATION: (3)
: OTHER INFORMATION: Xaa is Pro, Thr or Ser
: NAME/KEY: UNSURE
: LOCATION: (4)
: OTHER INFORMATION: Xaa is Leu, Ile, Val, Met, Ala or Pro
: NAME/KEY: UNSURE
: LOCATION: (5)
: OTHER INFORMATION: Xaa is any amino acid
: NAME/KEY: UNSURE
: LOCATION: (6)
: OTHER INFORMATION: Xaa is any amino acid
: NAME/KEY: UNSURE
: LOCATION: (7)
: OTHER INFORMATION: Xaa is Leu, Ile, Val, Met, Ala, Phe, Tyr or Trp
: NAME/KEY: UNSURE
: LOCATION: (8)
: OTHER INFORMATION: Xaa is Cys, Thr or Ser
: NAME/KEY: UNSURE
: LOCATION: (9)
: OTHER INFORMATION: Xaa is Arg, Lys or His
: NAME/KEY: UNSURE
: LOCATION: (10)
: OTHER INFORMATION: Xaa is any amino acid
: NAME/KEY: UNSURE
: LOCATION: (11)
: OTHER INFORMATION: Xaa is any amino acid
: NAME/KEY: UNSURE
: LOCATION: (12)
: OTHER INFORMATION: Xaa is Leu, Ile, Val, Met, Ala or Pro
: NAME/KEY: UNSURE

```

```

: LOCATION: (13)
: OTHER INFORMATION: Xaa is any amino acid
: NAME/KEY: UNSURE
: LOCATION: (14)
: OTHER INFORMATION: Xaa is any amino acid
: NAME/KEY: UNSURE
: LOCATION: (15)
: OTHER INFORMATION: Xaa is any amino acid
: NAME/KEY: UNSURE
: LOCATION: (16)
: OTHER INFORMATION: Xaa is Leu, Ile, Val, Met, Ala, Pro, Gly, Cys, Thr
: NAME/KEY: UNSURE
: LOCATION: (17) --(66)
: OTHER INFORMATION: Xaa can be any amino acid or no amino acid, position 17-65
: NAME/KEY: UNSURE
: LOCATION: (67)
: OTHER INFORMATION: Xaa is Leu, Ile, Val, Met, Ala or Pro
: NAME/KEY: UNSURE
: LOCATION: (68)
: OTHER INFORMATION: Xaa is any amino acid
: NAME/KEY: UNSURE
: LOCATION: (69)
: OTHER INFORMATION: Xaa is any amino acid
: NAME/KEY: UNSURE
: LOCATION: (70)
: OTHER INFORMATION: Xaa is Leu, Ile, Val, Met, Ala or Pro
: NAME/KEY: UNSURE
: LOCATION: (72)
: OTHER INFORMATION: Xaa is Leu, Ile, Val, Met, Ala, Pro or Gly
: NAME/KEY: UNSURE
: LOCATION: (73)
: OTHER INFORMATION: Xaa is Pro or Asn
: NAME/KEY: UNSURE
: LOCATION: (74) --(123)
: OTHER INFORMATION: Xaa can be any amino acid or no amino acid, position 74-123
: NAME/KEY: UNSURE
: LOCATION: (125) --(128)
: OTHER INFORMATION: Xaa is any amino acid
US-09-908-805B-59

Alignment Scores:
Pred. No.:      1.77e-06      Length:      128
Score:          123.00       Matches:      5
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches:      0
Query Match:    24.31%       Indels:         0
DB:             10          Gaps:         0

US-09-880-711-328 (1-350) x US-09-908-805B-59 (1-128)

Qy      185 NAGNAGNTGCGANN 171
      |||||:|||||:|||||
DB      68 *****TrpGly*** 72

RESULT 15
US-09-796-348A-26
: Sequence 26, Application US/09796338A
: Patent No. US20020061522A1
: GENERAL INFORMATION:
: APPLICANT: Millennium Pharmaceuticals, Inc.
: TITLE OF INVENTION: 1983, 52481, 2398, 45449, 50249, AND
: FILE REFERENCE: 52872, NVELL G PROTEIN-COUPLED RECEPTORS AND USES THEREFOR
: CURRENT APPLICATION NUMBER: US/09/796.338A
: PRIORITY FILING DATE: 2001-02-28
: PRIOR APPLICATION NUMBER: US 60/186,059
: PRIORITY FILING DATE: 2000-02-29

```

NUMBER OF SEQ ID NOS: 26
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO: 26
 LENGTH: 175
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: exemplary motif
 NAME/KEY: VARIANT
 LOCATION: 1-4, 6-53, 55-66, 68-137, 139-144, 146, 147, 150-170, 172,
 LOCATION: 173, 175
 OTHER INFORMATION: Xaa = any amino acid
 NAME/KEY: VARIANT
 LOCATION: 149
 OTHER INFORMATION: Xaa = any aromatic amino acid
 US-09-796-338A-26

Alignment Scores:
 Pred. No.: 1.75e+06 Length: 175
 Score: 123.00 Matches: 6
 Percent Similarity: 85.71% Conservative: 0
 Best Local Similarity: 85.71% Mismatches: 1
 Query Match: 24.31% Indels: 0
 DB: 10 Gaps: 0

US-09-880-711-328 (1-350) x US-09-796-338A-26 (1-175)

QY 190 TCCTGNAGNAGTTCANNNG 170
 |||||
 DB 145 CYS*****Gly***** 151

Search completed: April 28, 2003, 15:39:57
 Job time: 22 secs